

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:17:41 ; Search time 204 Seconds

(without alignments)
1988.427 Million cell updates/sec

Title: US-09-907-907a-42

Perfect score: 3557

Sequence: 1 DGFLPLRRDRALTOQLQVRA.....TAVLHNTQDNERINILLP 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3516	98.8	783	2	OBIX1
2	3512	98.7	783	2	OBIX1
3	3280	92.2	783	2	OBIX1
4	3033.5	85.3	748	2	OBIX1
5	2783	78.2	647	2	OBIX1
6	2389	67.2	540	2	OBIX1
7	2285	64.2	504	2	OBIX1
8	1962.5	55.2	784	2	OBIX1
9	1958	55.0	771	2	OBIX1
10	1940.5	54.6	720	2	OBIX1
11	1761	48.5	717	2	OBIX1
12	1365.5	38.4	991	2	OBIX1
13	1286.5	36.2	982	2	OBIX1
14	1276	35.9	745	2	OBIX1
15	1253.5	35.2	707	2	OBIX1
16	1212.5	34.1	760	2	OBIX1
17	1212.5	34.1	760	2	OBIX1
18	1204.5	33.9	703	2	OBIX1
19	1198	33.7	733	2	OBIX1
20	1193.5	33.6	698	2	OBIX1
21	1190.5	33.5	698	2	OBIX1
22	1188.5	33.4	698	2	OBIX1
23	1188.5	33.4	698	2	OBIX1
24	1188.5	33.4	698	2	OBIX1
25	1187.5	33.4	713	2	OBIX1
26	1187.5	33.4	713	2	OBIX1
27	1187.5	33.4	713	2	OBIX1
28	1187.5	33.4	713	2	OBIX1
29	1182.5	33.2	722	2	OBIX1
30	1182.5	33.2	722	2	OBIX1
31	1176	33.1	717	2	OBIX1

32	1175.5	33.0	713	2	Q7NGM0
33	1174	33.0	699	2	Q7P6G3
34	1172.5	33.0	705	2	Q895J3
35	1172.5	33.0	720	2	Q89WB3
36	1170	32.9	703	2	Q8RYA1
37	1169	32.9	715	2	Q8DGM9
38	1167.5	32.8	697	2	Q74CS9
39	1167.5	32.8	697	2	AAR34967
40	1166.5	32.8	713	2	Q7D283
41	1166.5	32.8	713	2	Q8U556
42	1165.5	32.8	701	2	Q8CST1
43	1164.5	32.7	697	2	Q6MMS2
44	1164.5	32.7	697	2	CAE79431
45	1163	32.7	717	2	Q8XXP6

ALIGNMENTS

RESULT 1

ID	OBIX1	PRELIMINARY:	PRT:	783 AA.
AC	OBIX1			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	3'-5' RNA exonuclease.			
GN	Name=OLD35;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed=1247348;			
RA	Leechzynecka M., Kang D.-C., Sarkar D., Su Z.-Z., Holmes M.,			
RA	Valerie K., Fisher P.B.;			
RT	"Identification and cloning of human polynucleotide phosphorylase,			
RT	hnpase (old-35), in the context of terminal differentiation and			
RT	cellular senescence."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16636-16641(2002).			
CC	-1-SIMILARITY: Contains 1 KH domain.			
DR	EMBL; AY027528; AAK13047.1; -.			
DR	HSSP; P05055; 1SRO.			
DR	GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.			
DR	GO; GO:0004527; F:exonuclease activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0006396; P:RNA processing; IEA.			
DR	InterPro; IPR001247; 3 Exonase.			
DR	InterPro; IPR001547; Glyco_hydro_5.			
DR	InterPro; IPR004087; KH.			
DR	InterPro; IPR004088; KH_type_1.			
DR	InterPro; IPR008994; Nucleic_acid_OR.			
DR	InterPro; IPR003029; SI.			
DR	Pfam; PF00013; KH_1; 1.			
DR	Pfam; PF03726; PNase; 1.			
DR	Pfam; PF01138; RNase; PH; 2.			
DR	Pfam; PF03725; RNase_PH_C; 2.			
DR	Pfam; PF00575; SI; 1.			
DR	SMART; SM00322; KH; 1.			
DR	SMART; SM00316; SI; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.			
DR	PROSITE; PS50084; KH_TYPE_1; 1.			
DR	PROSITE; PS50126; SI; 1.			
DR	Exonuclease.			
DR	SEQUENCE 783 AA; 85964 MW; 912939AF5309E2C CRC64;			

Query Match: 98.8%; Score 3516; DB 2; Length 783;

Best Local Similarity: 99.4%; Pred. No. 1.9e-201;

Matches 696; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

1 DGFLPLRRDRALTOQLQVRAVVDLGNKLEISSGKLARPDGSAVVOGSD 60

```

Db      18 DGPFLPRRRRALTQLOVRALMSSAGRAVAVLGNRKLEISSGKLARFADGSAVVOGSD 77
Qy      61 TAVVAVTAVSKTKPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
Db      78 TAVVAVTAVSKTKPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137
Qy      121 IRPLFPAGYFYDQVLCNLAVDGVNPDVLAINGASVALSLSDIPNNGPGVAVRIGIID 180
Db      138 IRPLFPAGYFYDQVLCNLAVDGVNPDVLAINGASVALSLSDIPNNGPGVAVRIGIID 197
Qy      181 GEYVAVNPTRKEMSSSTLNLVAVGAPKSOIYVLEASAEINIIQODFCHAIKVGKVTQOIIQ 240
Db      198 GEYVAVNPTRKEMSSSTLNLVAVGAPKSOIYVLEASAEINIIQODFCHAIKVGKVTQOIIQ 257
Qy      241 GIQOLVETGVTYKTPOKLFTPSPEIYKTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db      258 GIQOLVETGVTYKTPOKLFTPSPEIYKTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 317
Qy      301 LDTEEOUKERPEADPYEIIIESFNVVAKVFRSIVLNEYKRCDCRDLTSLNVCCEYDMF 360
Db      318 LDTEEOUKERPEADPYEIIIESFNVVAKVFRSIVLNEYKRCDCRDLTSLNVCCEYDMF 377
Qy      361 KTLHGSALFORGOVQVCTVTFDSLSGSKSDOYITAINGIKDKNFMHYEPFYATNEI 420
Db      378 KTLHGSALFORGOVQVCTVTFDSLSGSKSDOYITAINGIKDKNFMHYEPFYATNEI 437
Qy      421 GKVTGLNRRELGHGALAEKALPYIPRDPFTTIRVTSEVLESSSSMASACGSLAMD 480
Db      438 GKVTGLNRRELGHGALAEKALPYIPRDPFTTIRVTSEVLESSSSMASACGSLAMD 497
Qy      481 SGVIVSSAVAGVALGYTKTPDEKGEIEDYRLLDIIGIEBYNDMDPKTAGTKGTAL 540
Db      498 SGVIVSSAVAGVALGYTKTPDEKGEIEDYRLLDIIGIEBYNDMDPKTAGTKGTAL 557
Qy      541 QADIKLKGIPKIKYWEAIIQASVAKKEIIQMNKTIKSPRASRKENGVPVETVOVPSKR 600
Db      558 QADIKLKGIPKIKYWEAIIQASVAKKEIIQMNKTIKSPRASRKENGVPVETVOVPSKR 617
Qy      601 AKFVPGGVNKLKQIQAETGVTTISQVDEBTSVFAPTSPVMEHADFTEICKDQEOOLE 660
Db      618 AKFVPGGVNKLKQIQAETGVTTISQVDEBTSVFAPTSPVMEHADFTEICKDQEOOLE 677
Qy      661 FGAVYVATITIRPTGVNWKIYPMNTAVLNLNTOLDNERL 700
Db      678 FGAVYVATITIRPTGVNWKIYPMNTAVLNLNTOLDNERL 717

RESULT 2
ID      Q8TCS8      PRELIMINARY;      PRT;      783 AA.
AC      Q8TCS8;
DT      01-JUN-2002 (TRENBLREL 21, Created)
DT      01-JUN-2002 (TRENBLREL 21, Last sequence update)
DT      01-OCT-2004 (TRENBLREL 28, Last annotation update)
DE      Polynucleotide phosphorylase-like protein (EC 2.7.7.8)
DE      (Polynucleotide nucleotidyltransferase 1).
GN      Homo sapiens (Human).
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=22307406; Pubmed=12419256;
RA      Rajmakers R., Vree Egberts W., van Venrooij W., Pruijn G.;
RT      "Protein-protein interactions between human exosome components support
RT      the assembly of RNase PH-type subunits into a six-membered PNPase-like
RT      ring."
RL      J. Mol. Biol. 323:653-663 (2002).
RC      [2]
RP      SEQUENCE FROM N.A.
RX      TISSUE=Skin;
RX      MEDLINE=22386257; Pubmed=12477932;

```

```

RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marisica K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udén T.B., Toehiyuki S., Cantirci P., Prange C.,
RA      Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA      Krzywinski M.I., Skalka U., Smallus D.E., Schnerch A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences."
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RA      Strausberg R.;
RL      Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC      -1- Similarity: Contains 1 KH domain.
DR      EMBL; AJ458465; CAD30289.1; -.
DR      EMBL; BC053660; AAH53660.1; -.
DR      PIR; T50626; T50626.
DR      HSSP; P05055; ISRO-.
DR      GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR      GO; GO:0004654; F:polynucleotide nucleotidyltransferase a. .; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006396; P:RNA processing; IEA.
DR      InterPro; IPR001247; 3 Exonase.
DR      InterPro; IPR001547; Glyco_hydro_5.
DR      InterPro; IPR004087; KH.
DR      InterPro; IPR004088; KH_type_1.
DR      InterPro; IPR008994; Nucleic_acid_ob.
DR      InterPro; IPR003029; S1.
DR      Pfam; PF00013; KH 1; 1.
DR      Pfam; PF03726; PNPase; 1.
DR      Pfam; PF01138; RNase PH; 2.
DR      Pfam; PF03725; RNase PH_C; 2.
DR      Pfam; PF00575; S1; 1.
DR      SMART; SM00322; KH; 1.
DR      SMART; SM00316; S1; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR      PROSITE; PS50084; KH_TYPE_1; 1.
DR      PROSITE; PS50126; S1; 1.
KW      Nucleotidyltransferase; Transferase.
SQ      SEQUENCE 783 AA; 85936 MW; 8A3629AF52F8E24 CRC64;

Query Match      98.7%; Score 3512; DB 2; Length 783;
Best Local Similarity 99.3%; Pred. No. 3.4e-201;
Matches 695; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 241 GIGOLVKEGTGKATPOKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
 DB 258 GIGOLVKEGTGKATPOKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317
 QY 301 LDTBEOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEVDMF 360
 DB 318 LDTBEOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEVDMF 377
 QY 361 KTLHGSALFORGOVQVCTVTPDSLESIGSDQVITAINGIKDNPMFLHYEPFYATNEI 420
 DB 378 KTLHGSALFORGOVQVCTVTPDSLESIGSDQVITAINGIKDNPMFLHYEPFYATNEI 437
 QY 421 GKVTGLRRELGHGALAEKALPYIPDPFPTITVTSSEVLESNGSSSMASACGSLALMD 480
 DB 438 GKVTGLRRELGHGALAEKALPYIPDPFPTITVTSSEVLESNGSSSMASACGSLALMD 497
 QY 481 SGVPISSAVAGVALGVTCTDPEKGEIEDVRLTDIGIEDYNGDMDFKTAGNKGITLAL 540
 DB 498 SGVPISSAVAGVALGVTCTDPEKGEIEDVRLTDIGIEDYNGDMDFKTAGNKGITLAL 557
 QY 541 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTTISKPRASRENGPVVETVOVPLSKR 600
 DB 558 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTTISKPRASRENGPVVETVOVPLSKR 617
 QY 601 AKFVPGPGYNLKKLQAGTGVITISOVDEETSFVPAFTPSVMEADFTIEICKDOEOOLE 660
 DB 618 AKFVPGPGYNLKKLQAGTGVITISOVDEETSFVPAFTPSVMEADFTIEICKDOEOOLE 677
 QY 661 FGAVYATITEIRDTGVWVVKLPYMTAVLHNTOLDNERL 700
 DB 678 FGAVYATITEIRDTGVWVVKLPYMTAVLHNTOLDNERL 717

RESULT 3

Q8KIR3 PRELIMINARY; PRT; 783 AA.
 AC Q8KIR3; ID Q8KIR3; AC Q8KIR3; DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2004 (Tremblrel. 28, Last annotation update)
 DE Polynucleotide phosphorylase (EC 2.7.7.8).
 GN Name=Pnpl1; Synonym=pnpase;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN RP
 RP SEQUENCE FROM N.A.
 RC STRAIN=divers;
 RX MEDLINE=22307406; PubMed=12419256;
 RA Rajmankers R., Vree Egberts W., van Venrooij W., Pruijn G.;
 RT "Protein-protein interactions between human exosome components support
 RT the assembly of RNAse PH-type subunits into a six-membered PNPase-like
 RT ring";
 RT J. Mol. Biol. 323:653-663(2002).
 RN RP
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhac N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stetler-Munroe M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinci P., Prange C.,
 RA Rata S.S., Loggellano N.A., Peters G.J., Abraham R.D., Mullany S.J.,
 RA Boek S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodríguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,
 RA Krzywnicki M.I., Skalka U., Smalms D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN RP
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RA Strauberg R.;
 RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- Similarity: Contains 1 KH domain.
 DR EMBL; AJ507387; CAD45436.1; -;
 DR EMBL; BC055826; AAH55826.1; -;
 DR HSSP; P05055; ISRO.
 DR MD; MG1.1918951; Pnpl1.
 DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
 DR GO; GO:0004654; F:Polyribonucleotide nucleotidyltransferase a. . .; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006396; P:RNA processing; IEA.
 DR InterPro; IPR001247; 3-ExoRNase.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR InterPro; IPR004087; KH.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR008994; Nucleic_acid_DB.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF00013; KH_1; 1.
 DR Pfam; PF03726; PNPase; 1.
 DR Pfam; PF01138; RNase_H; 2.
 DR Pfam; PF03725; RNase_H; 2.
 DR Pfam; PF00575; S1; 1.
 DR SMART; SM00322; KH; 1.
 DR SMART; SM00316; S1; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PROSITE; PS50084; KH_TYPE_1; 1.
 DR PROSITE; PS50126; S1; 1.
 KW Nucleotidyltransferase; Transferase.
 SQ SEQUENCE 783 AA; 85682 MW; F35FEB91AAB5626 CRC64;

Query Match 92.28; Score 3280; DB 2; Length 783;
 Best local similarity 91.08; Pred. No. 2.5e-187;
 Matches 636; Conservative 39; Mismatches 24; Indels 0; Gaps 0;

QY 2 GPRLPRDADLTOLQVRALMSSAGSAVAVDGNRLKLEISGKGLAPADGCAVIGSDT 61
 DB 19 GPLGRPGNRNALSTLQKRALMSSGSAVAVDGLHRLKLEISGKGLAPADGCAVIGSDT 78
 QY 62 AVMTAVSKTKPSPQMPVVDYRQKAAAGRIPTYLRREVSTDEILTSRIIDRSI 121
 DB 79 AVMTAVSKTKASQMPVVDYRQKAAAGRIPTYLRREVSTDEILTSRIIDRSI 138
 QY 122 RPLPAGYFYDTQVLCNLAVDGNBPDVLAINGASVALSLDSIPMNGPVAVRIGIIDG 181
 DB 139 RPLPAGYFYDTQVLCNLAVDGNBPDVLAINGASVALSLDSIPMNGPVAVRIGIIDG 198
 QY 182 EYVNPTRKEMSSSTLNLVVAAGPKSQIWLLESAENILODPCHAIKGVKTYQOIIIG 241
 DB 199 ECVNPPRRREMSSTLNLVVAAGPKSQIWLLESAENILODPCHAIKGVKTYQOIIIG 258
 QY 242 IQOLVKEGTGKATPOKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 301
 DB 259 IQOLVKEGTGKATPOKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 318
 QY 302 DTEBOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEVDMF 361
 DB 319 DTEBOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEVDMF 378
 QY 362 TLHGSALFORGOVQVCTVTPDSLESIGSDQVITAINGIKDNPMFLHYEPFYATNEIG 421
 DB 379 TLHGSALFORGOVQVCTVTPDSLESIGSDQVITAINGIKDNPMFLHYEPFYATNEIG 438

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smalley D.E., Scherch A., Schein J.E.,
 RA Jones S.U., Warr M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RC TISSUE=Mammary tumor. Metalloprotein-TGF alpha model. 10 month old
 RC virgin mouse. Taken by biopsy.;
 RA Strauberg R.;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 1 KH domain.
 DR EMBL: BC027228; AAR27228.2; -.
 DR MGI: MGI:1918951; Pppl1.
 DR GO: GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0006396; P:RNA processing; IEA.
 DR InterPro: IPR001247; 3_ExoRNase.
 DR InterPro: IPR001547; Glyco_hydro_5.
 DR InterPro: IPR004088; KH_type_1.
 DR InterPro: IPR008994; Nucleic_acid_OB.
 DR InterPro: IPR003029; S1.
 DR Pfam: PF000013; KH_1; 1.
 DR Pfam: PF03726; NPase; 1.
 DR Pfam: PF01138; RNase_PH; 2.
 DR Pfam: PF03725; RNase_PH_C; 2.
 DR Pfam: PF00575; S1; 1.
 DR SMART: SM00332; KH; 1.
 DR SMART: SM00316; S1; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PROSITE: PS50084; KH_TYPE_1; 1.
 DR PROSITE: PS50126; S1; 1.
 DR NON TER 1 1
 SQ SEQUENCE 647 AA; 7105 MW; 5B1574DBDA723B43 CRC64;
 Query Match 78.2%; Score 2783; DB 2; Length 647;
 Best Local Similarity 92.4%; Pred. No. 9.2e-156;
 Matches 537; Conservative 29; Mismatches 15; Indels 0; Gaps 0;
 QY 120 SIRLPAGYPTDQVLCNLAVDGNVDPDLANGSVALSLSDIPNGVGVARRIGI 179
 DB 1 SIRLPAGYPTDQVLCNLAVDGNVDPDLANGSVALSLSDIPNGVGVARRIGI 60
 QY 180 DGEVYVNPTRKEMSSSTLNLVAVAGPKSQIYMLEASANIQQDFCHAIKGVKYYTOOI 239
 DB 61 DGEVYVNPTRKEMSSSTLNLVAVAGPKSQIYMLEASANIQQDFCHAIKGVKYYTOOI 120
 QY 240 QGIQQLVKEGTGVTTRTPQKFTPSPEIYKYTHKLAMERLVAVFTDYEDHKVSRDEAVNKI 299
 DB 121 QGIQQLVKEGTGVTTRTPQKFTPSPEIYKYTHKLAMERLVAVFTDYEDHKVSRDEAVNKI 180
 QY 300 RLDTREBOLKEFPADPEIETESFNVVAKVFRSIVLNEYKRCGGRDITSLRANVSCEDVM 359
 DB 181 RLDTREBOLKEFPADPEIETESFNVVAKVFRSIVLNEYKRCGGRDITSLRANVSCEDVM 240
 QY 360 PFTLHSGALFORGTOVLCVTPDLSLESGISDQVITANGIKDKNFMLATPEPPYATNE 419
 DB 241 PFTLHSGALFORGTOVLCVTPDLSLESGISDQVITANGIKDKNFMLATPEPPYATNE 300
 QY 420 IGVKTGNRLRELHGALAEKALCVPIPRDPFTIRVTSVLESNGSSMASACGSLALM 479
 DB 301 TGVKTGNRLRELHGALAEKALCVPIPRDPFTIRVTSVLESNGSSMASACGSLALM 360
 QY 480 DSGVPISSAVAGVAVIGVTKDPEKGEIEDRLTDLIGIDYNGDMFKATGNKGTGA 539
 DB 361 DSGVPISSAVAGVAVIGVTKDPEKGEIEDRLTDLIGIDYNGDMFKATGNKGTGA 420
 QY 540 LQADIKLPGVIFIKIWEAIQOASVAKKEIIQIMNKTTISKPRASREKNGPVETVQVPLSK 599

DB 421 LQADIKLPGVIFIKIWEAIQOASVAKKEIIQIMNKTTISKPRASREKNGPVETVQVPLSK 480
 QY 600 RAKFVGGGVNKLQLEAGTTSISQVDEFTSVAPRPSVWHEARDPTTEICXDDQEQOL 659
 DB 481 RAKFVGGGVNKLQLEAGTTSISQVDEFTSVAPRPSVWHEARDPTTEICXDDQEQOL 540
 QY 660 EFGAVYATITTEIRDGTGVNVLVYPMNTAVLLHNTQNLNERL 700
 DB 541 EFGAVYATITTEIRDGTGVNVLVYPMNTAVLLHNTQNLNERL 581
 RESULT 6
 ID Q9DC52 PRELIMINARY; PRT; 540 AA.
 AC Q9DC52.
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male lung cDNA, RIKEN full-length enriched library,
 DE clone:120003F12 product:hypothetical 3' exoribonuclease family
 DE containing protein, full insert sequence.
 GN Name=Prpct1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RA The PANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Mizawa K., Nagaoke S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitunai T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara B., Matsumaki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]

RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=Lung;
 RA Aachai J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arawaka T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuo M.,
 RA Hanagaki T., Hara A., Hayashi N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai U., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sojabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK004563; BAB2374.1; -
 DR HSSP: Q53597; 1R3P.
 DR MGD: MGI:1918951; Pnpf1.
 DR GO: GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0006396; P:RNA processing; IEA.
 DR InterPro: IPR001247; 3 ExoRNase.
 DR InterPro: IPR001547; Glyco_hydro_5.
 DR Pfam: PF03726; PNPase; 1.
 DR Pfam: PF01138; RNase_PH; 2.
 DR Pfam: PF03725; RNase_PH_C; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KM Hypothetical protein.
 SQ SEQUENCE 540 AA; 58938 MW; 457BFA3E3579A072 CRC64;

Query Match 67.2%; Score 2389; DB 2; Length 540;
 Best Local Similarity 90.1%; Pred. No. 2.5e-134;
 Matches 465; Conservative 29; Mismatches 22; Indels 0; Gaps 0;

QY 2 GPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSDT 61
 DB 19 GFLRPGRNRLSTLOVRALWSSAGSRAVYVDLGNRKLEISSGKLARFADGSAVVGSDT 78
 QY 62 AAVTAVSKTKPSBSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSI 121
 DB 79 AAVTAVSKTKPSBSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSI 138
 QY 122 RPLFPAGFYDTQVLCNLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIIDG 181
 DB 139 RPLFPAGFYDTQVLCNLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIIDG 198
 QY 182 EYVNPTRKEMSSSTLNIVAGAPKSOIVMLEASAEINLODPFGHAIKVGKTYQOIIQG 241
 DB 199 ECVNPFRRKEMSSSTLNIVAGAPKSOIVMLEASAEINLODPFGHAIKVGKTYQOIIQG 258
 QY 242 IQOLVKEGTGVRTPQKLFTPSPRIKVTYHKLAMERLVAFTDYEHDKVSHDAVNKIRL 301
 DB 259 IQOLVKEGTGVRTPQKLFTPSPRIKVTYHKLAMERLVAFTDYEHDKVSHDAVNKIRL 318
 QY 302 DTEBQLEKPEADPYEIIIESFNVAKEVPSIYLVNEYKRCGDLTSLRVNSECVDMEK 361
 DB 319 DTEBQLEKPEADPYEIIIESFNVAKEVPSIYLVNEYKRCGDLTSLRVNSECVDMEK 378
 QY 362 TLHGSALEFORGOVQLCTVTFPSLESIGKSDOVITAINGIDKPKMLHYEPPTATNEIG 421
 DB 379 TLHGSALEFORGOVQLCTVTFPSLESIGKSDOVITAINGIDKPKMLHYEPPTATNEIG 438
 QY 422 KVTGLNRELGHGALAEKALVPIPRDFFTIKRTSEVLESNGSSVMAACGGSIALMDS 481
 DB 439 KVTGLNRELGHGALAEKALVPIPRDFFTIKRTSEVLESNGSSVMAACGGSIALMDS 498
 QY 482 GVPISSAVAGVAGLVTKTDEKGIEDYRLITDIL 517
 DB 499 GVPISSAVAGVAGLVTKTDEKGIEDYRLITDIL 534

RESULT 7
 Q96T05 PRELIMINARY; PRT; 504 AA.
 Q96T05

AC Q96T05;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ14531.
 OS Homo sapiens (Human)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14702039;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Niimomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hociuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiyia S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Mitsuashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshioka K., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Wakebe H.,
 RA Nakagawa S., Senoh A., Mizouchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi Y., Fujita Y.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohnori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs";
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL: AK027437; BAB55109.1; -
 DR HSSP: Q53597; 1R3H.
 DR GO: GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006396; P:RNA processing; IEA.
 DR InterPro: IPR001247; 3 ExoRNase.
 DR InterPro: IPR001547; Glyco_hydro_5.
 DR Pfam: PF03726; PNPase; 1.
 DR Pfam: PF01138; RNase_PH; 2.
 DR Pfam: PF03725; RNase_PH_C; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KM Transferase.
 SQ SEQUENCE 504 AA; 55998 MW; 2BB89ADB409322D6 CRC64;

Query Match 64.2%; Score 2285; DB 2; Length 504;
 Best Local Similarity 100.0%; Pred. No. 3.7e-128;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSD 60
 DB 18 DGPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSD 77
 QY 61 TAAVAVSKTKPSBSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRS 120
 DB 78 TAAVAVSKTKPSBSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRS 137
 QY 121 IRPLFPAGFYDTQVLCNLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
 DB 138 IRPLFPAGFYDTQVLCNLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIID 197

```

Qy 181 GYYVNPTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 240
Db 198 GEYVNPTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 257
Qy 241 GIGQLVKETGVTKRTPKQKFTPSPEIYKTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db 258 GIGQLVKETGVTKRTPKQKFTPSPEIYKTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317
Qy 301 LDTEBOLKEKPPADPEIIESFNVAKEVRSIVANEYKCCDGRDLTSLRNVSCEDVDF 360
Db 318 LDTEBOLKEKPPADPEIIESFNVAKEVRSIVANEYKCCDGRDLTSLRNVSCEDVDF 377
Qy 361 KTLHGSALFORGOTVCTVTFPSLESIGKSDOYITAINIKDKNFMPLYEPYATNEI 420
Db 378 KTLHGSALFORGOTVCTVTFPSLESIGKSDOYITAINIKDKNFMPLYEPYATNEI 437
Qy 421 GKVTGLNRRELGHGALAEKALYPYIPRDPF 450
Db 438 GKVTGLNRRELGHGALAEKALYPYIPRDPF 467

RESULT 8
ID 070228 PRELIMINARY; PRT; 784 AA.
AC 070228:
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE ACP9005.
GN Name=agCG49269; ORFNames=ENSANG00000018771;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1 CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAB01068979; EMBL3718.1;
DR GO: GO:0000175; F:3'-5'-exonibonuclease activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0006396; P:RNA processing; IEA.
DR InterPro: IPR001247; 3 ExonName.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR003029; SL.
DR Pfam: PF00013; KH_1; 1.
DR Pfam: PF03726; ENPase; 1.
DR Pfam: PF01138; RNase PH; 2.
DR Pfam: PF03725; RNase PH; 2.
DR PROSITE: PSS0084; KH_type_1; 1.
DR PROSITE: PSS0125; SL_1.
SQ SEQUENCE 784 AA; 86050 MW; D09241BF3A1C4A30 CRC64;

Query Match 55.2%; Score 1962.5; DB 2; Length 784;
Best Local Similarity 55.4%; Pred. No. 1,3e-108;
Matches 387; Conservative 119; Mismatches 187; Indels 5; Gaps 4;

```

```

Qy 184 VNPTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 243
Db 195 IINPTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 254
Qy 244 QLVKGVTKRTPKQKFTPSPEIYKTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLOT 303
Db 255 RLQGVKGVTKRTPKQKFTPSPEIYKTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLOT 314
Qy 304 BEOLKEKPPADPEIIESFNVAKEVRSIVANEYKCCDGRDLTSLRNVSCEDVDF 363
Db 315 IDKWSFPFADPEIIESFNVAKEVRSIVANEYKCCDGRDLTSLRNVSCEDVDF 374
Qy 364 HGSALFORGOTVCTVTFPSLESIGKSDOYITAINIKDKNFMPLYEPYATNEI 423
Db 375 HGSALFORGOTVCTVTFPSLESIGKSDOYITAINIKDKNFMPLYEPYATNEI 434
Qy 424 TGLNRRELGHGALAEKALYPYIPRDPFTIRVTSVLENSGSSMSACGSLALMDGV 483
Db 435 GPGRRELGHGALAEKALYPYIPRDPFTIRVTSVLENSGSSMSACGSLALMDGV 494
Qy 484 PISAVAGVAGVTK-TDEKGBIEDYRLTLGIEDYNGMDPIAGTNGITLQA 542
Db 495 PVCEAAAGVAGVITKXENNDTKHLDYCLTLGLIEDYNGMDPIAGTNGITLQA 554
Qy 543 DTKLPGIPKIVMEAIQASVAKKEIIOINKTKISKRPAKENGVPVETVOVPLSKRAK 602
Db 555 DTKVPGIPKIVMEAIQASVAKKEIIOINKTKISKRPAKENGVPVETVOVPLSKRAK 614
Qy 603 FVPGGYNLKRLQAEVTTSIOVDEETSVAPPTPSVMEHARDPTEICDDOEOLDFG 662
Db 615 LFGGGTGLNRRLVLETVGLTPDDETSFRIFAPSEAMHKEYLENLAKEKIPDLEFG 674
Qy 663 AVTTATTEIRDTGVKYLKYPMTAVILNTOLENEU 700
Db 675 AITVARIVELRDTGVWTLVPSMPTLHNSOLDORXI 712

RESULT 9
ID 09V9X7 PRELIMINARY; PRT; 771 AA.
AC 09V9X7:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CG11337-PA (CG11337-pb).
GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Buttle K.C., Buzan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertile S., Fleischmann W.,
RA Foster C., Gabrielle A.E., Garcia N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harrie N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```


RA Hostali D., Housen F.A., Howland T.J., Wei M.H., Ibeagwam C.,
RA Jatalin M., Kalush K., Kapen G.H., Ke Z., Kemnison J.A., Kechem K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao X., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paley J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodager, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RA Science 287:2185-2195(2000).
RL [2]
RN SEQUENCE FROM N.A.
RN MEDLINE=22426065; PubMed=12537568;
RP Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Swirskas R.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Paley J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Swirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RA "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RA *melanogaster* euchromatic genome sequence";
RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RL [3]
RN SEQUENCE FROM N.A.
RN MEDLINE=22426070; PubMed=12537573;
RP Kaminler J.S., Bergman C.M., Krommiller B., Carlson J., Swirskas R.,
RA Kaminler J.S., Bergman C.M., Krommiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the *Drosophila melanogaster* euchromatin:
RA a genomic perspective.";
RA Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RL [4]
RN SEQUENCE FROM N.A.
RN MEDLINE=22426069; PubMed=12537572;
RP Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminler J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RA systematic review";
RA Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RL [5]
RN SEQUENCE FROM N.A.
RN FLYBASE;
RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RP FLYBASE;
CC Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 KH domain.
DR EMBL; AEO03778; AAF57151.2; -.
DR HSSP; O53597; IE3P.
DR InAcct; Q9V9X7; -.
DR FLYBASE; FBGN0039846; CG11337.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3 Exonase.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH_type_1.

[illegible]


```

GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
NCBI_TaxID=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorese V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,
RA Munoz J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Cejnkler S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 KH domain.
DR EMBL; AY061061; AAL28609.1; -.
DR HSSP; Q53597; 1E3P.
DR FLYBASE; FBgn0039846; CG11337.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3_ExonNase.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH type 1.
DR InterPro; IPR008994; Nucleic_acid_ob.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH_1; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PSS0084; KH_TYPE_1; 1.
DR PROSITE; PSS0126; S1; 1.
DR SEQUENCE 720 AA; 79381 MW; DD54952528ESDBDF CRC64;

Query March 54.6%; Score 1940.5; DB 2; Length 720;
Best Local Similarity 56.2%; Pred. No. 2.4e-107;
Matches 374; Conservative 123; Mismatches 164; Indels 5; Gaps 4;

QY 39 LEISSGLKARPADSSAVVQSGDTAVMTVATSKTSPSQ-EMPVVDYRQKAAAGRIPT 97
DB 1 MTFSSGLRARPANGTAVQCMGDTAVMTVATVAKAKPNPQGFMPVVDYRLKNAASGRIPM 60

QY 98 NYLREVGTSKXELITSRIIRSRIRPLFPAGYPTDTQVLCMLLAVDGVNEVDVLAINGAS 157
DB 61 NFMRELGSPSEKELISARLIDRSIRPLFKDYRTETQVLCMLLAVDGVNEVDVLAINGAS 120

QY 158 VALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVYAGAPKQIWLKASAE 217
DB 121 MALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVYAGAPKQIWLKASAE 180

QY 218 NILQDFCHALVGVKVTQOIIQGIQOLVKETGVTKTPQKLTSPSEIVYVTKLAMEK 277
DB 181 VVLQDDLLKAIKQGTREAOPIIHEIBRLQKAYGKQKEVEVAADVELGAVASWCEMR 240

QY 278 LYAVFTYEHDKVSRDEAVNKIRLDTEBQLEKFPPEADPYEIIISFNVAKEVRSVLYN 337
DB 241 LREIFQOSTHDKMSRDNAVNEVRSNVIDKWSFPDEPESLITQFQFQTSITIRRELIPE 300

QY 338 EYKCDGRDLTSLNVSCGVDMPTKLHGSALEFGQGVQLCTVFPDSLESGIKSDQVITA 397
DB 301 RGLRCDGRDIQOLNINISQVDMKPLHGSALEFGQGVQLCTVFPDSLESGIKSDQVITA 359

QY 398 IN--GIQDKNFMLYBEPYVATNEIGKVTGLRRRLGHGALAKALYVPIPRDEPPIRV 455
DB 360 LDSGLKAKNFMLYBEPYVATNEIGKVTGLRRRLGHGALAKALYVPIPRDEPPIRV 419

QY 456 TSEVLSENGSSMASACGSLALMDSGVPISSAVAGVATGVTYK--TDEPKKEIDYRLLT 514
DB 420 TSEVLSENGSSMASACGSLALMDSGVPISSAVAGVATGVTYK--TDEPKKEIDYRLLT 479

```

```

QY 515 DILGIEDYNDGMDKINAGTKNGITATLQADIKLPKIPKIVMEALIQASVAKKEILQIMNK 574
DB 480 DILGIEDYNDGMDKINAGTKNGITATLQADIKLPKIPKIVMEALIQASVAKKEILQIMNK 539

QY 575 TISKPRASRKXNGVYVETVOVPLSKAKFVPGPGYINLKKLOAETGVITISQVDETSVPA 634
DB 540 AIREPRRYKPSWSESTLTYVEPOORALIGPSGLHMKRIYLETGSTLTAVDEFHVFVA 599

QY 635 PTPSVMEARDPFIETICKDQOEQLEFGAVYATITETIRDTGVWVKKYPMNTAVTLANTQ 694
DB 600 PSCAMMEARKELEGVWYKERVDPLEFGITAKITELRDTGVWVILYPSMPALLHNSQ 659

QY 695 LDNERL 700
DB 660 LDQKRI 665

RESULT 11
Q8IH29 PRELIMINARY; PRT; 717 AA.
ID Q8IH29;
AC Q8IH29;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DB GMI6802P (Fragment).
GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
NCBI_TaxID=7227;
RP SEQUENCE FROM N.A.
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorese V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Lao G.,
RA Miranda A., Mungall C.J., Munoz J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Cejnkler S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 KH domain.
DR EMBL; BT001457; AAN71212.1; -.
DR FLYBASE; FBgn0039846; CG11337.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3_ExonNase.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH type 1.
DR InterPro; IPR008994; Nucleic_acid_ob.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH_1; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PSS0084; KH_TYPE_1; 1.
DR PROSITE; PSS0126; S1; 1.
DR NON TER 1
SQ SEQUENCE 717 AA; 79403 MW; 6C1CA7ABC1714D02 CRC64;

Query March 49.5%; Score 1761; DB 2; Length 717;
Best Local Similarity 54.4%; Pred. No. 1.3e-96;
Matches 336; Conservative 119; Mismatches 159; Indels 4; Gaps 3;

QY 86 ROKAAAAGRIPTVYLRREVSTSKELITSRIIRSRIRPLFPAGYPTDTQVLCMLLAVDGV 145
DB 46 RMTFSSGRIIPMNFMRRELGPSEKELISARLIDRSIRPLFKDYRTETQVLCMLLAVDGV 105

QY 146 NEPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVYAGAP 205
DB 106 HSPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVYAGAP 165

```



```
ID OGA10 PRELIMINARY; PRT; 982 AA.
AC OGA10;
DT 05-JUN-2004 (TEMBLrel. 27, Created)
DT 05-JUN-2004 (TEMBLrel. 27, Last sequence update)
DE Putative polyribonucleotide nucleotidyltransferase.
GN Name-OJ1014_H03.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003980; BAD21450.1; -
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001247; 3 ExoRNase.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR001029; S1.
DR Pfam; PF03726; RNase; 1.
DR Pfam; PF01138; RNase; PH; 2.
DR Pfam; PF03725; RNase; PH; 2.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 2.
DR PROSITE; PS50126; S1; 1.
DR Transferrase.
SQ SEQUENCE 982 AA; 107027 MW; 0B05E5A0E8684351 CRC64;

Query Match 36.2%; Score 1286.5; DB 2; Length 982;
Best Local Similarity 39.3%; Pred. No. 4.5e-68;
Matches 286; Conservative 138; Mismatches 253; Indels 51; Gaps 13;

QY 8 RRDALTOLOV---RALMSAGSRAVA-----VDLGNKLEISS 43
DB 17 RRARFPAPLSPVGRAPFLSGAAAEVAQADAPPPPPGKRVLESFREFEIGGRVISFET 76
QY 44 GKLAEPDGSAAVVGSGTAVMT-AVSKTKPSQPMPLVVDYRQKAAAAGRIPTNYLR 102
DB 77 GKMAFPANGSVVISMDDTHVLSVAAKSSSEPRDFPLVVDYQEKQYAGVIPPYMR 136
QY 103 EVGSDKEILTSRIIDRSIRPLPAGYFYDTQV-----LCNLAVGVNPEVLAINGA 156
DB 137 EGAPKERELLCGRILIDRPIPLPPPGFYHEVQVNMATIIWNVSISSGKODPPVMAANAS 196
QY 157 SVSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNVVAGAPKSOIWMLEASA 216
DB 197 SAALMLSDIPWNGIIGVIRGRIDGNFVLANPTVDELGSLDLNVYA-CSRDKTLMDVQA 255
QY 217 ENLIQDFCHAIKVGKYYTOOIIQGIQOLVKETGVTKTPQKLFPPSPBEIVKTHKLAME 276
DB 256 REITERDLQGMKLAHAENAKCINPOLRLAKRAG-KKKKEKYKISLSDSKYKIRTLSEA 314
QY 277 RLVAFTDYHDKVSDEAVNKIRLDTBEOLKKEKFPADPYEIIIESFNVAKEVFRSIVL 336
DB 315 PIEEVFTDSTYGKEFGEALENITQSVAKLBECCDESLKFLHKAVDYTRKQVIRKRII 374
QY 337 NEYKRCGRDLTSLRNVSCVDMEFKTLHGSALFORGOTVLCVTPDSLESIGISD-QVI 395
DB 375 EKGIRVGRQLDREVRLPYCCSSSTYPIHGSALFSRGDTQVCTYTL-----GAGDQRL 429
QY 396 TAINGIKDKFMFLHYEPPPYATNEIGKVTGLNRRELHGALAEKALYVPVIR-RDPPTI 453
DB 430 DSIIGPPTKRFMLHYSPFPFSINEVAKRGGLNRREVGHTLAERKALLAVLPPEGEFPPTV 489
QY 454 RYSEVLESNSSSSMASCGSGLALMDSGVPISSAVGVALGVTNKDPEKGELEDRL 513
DB 490 RVNSEVWASDSTSMASVCGSSMALMDGIVRHHVAVSVGLVSEVDQTTGDISSYRIL 549
QY 514 TDIIGIEDYNDMDPKTAGTKGITALQADIKLPGIPIKIWEAIIQASVAKKEILQIMN 573
DB 550 TDIIGLEDHLDGMDPKTAGTRGITATQDLKPKAGIPLDITCEGLERPKARKRNQILDRMD 609
```

```
QY 574 KTISKPRASRKEKNGPVETQVPLSKRAKFGVPGYNLKKQQAETGVITISQVDEFTFSVF 633
DB 610 QEISSARAFVNDSSPPLATLISFSSDSLRLKL-----PFRKKIEGETGARVS-VSDGVITIV 664
QY 634 APFSPVWEHARDFITTEICDKDDEQQLSEFGAVYATTTETRDYGVWYKILPMTAVULHNT 653
DB 665 AKTQPIMDKAIKEKVEFLV-----GREIEVGRYKGVSVSIKEYGAFVEFNGQOG-LIHIS 719
QY 694 QLNERLUN 701
DB 720 ELSHDKVS 727

RESULT 14
ID BAD21450 PRELIMINARY; PRT; 982 AA.
AC BAD21450;
DT 01-JUN-2004 (TEMBLrel. 27, Created)
DT 01-JUN-2004 (TEMBLrel. 27, Last sequence update)
DE Putative polyribonucleotide nucleotidyltransferase.
GN OJ1014_H03.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza; Oryza sativa.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone OJ1014_H03.14";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003980; BAD21450.1; -
DR Transferrase.
SQ SEQUENCE 982 AA; 107027 MW; 0B05E5A0E8684351 CRC64;

Query Match 36.2%; Score 1286.5; DB 2; Length 982;
Best Local Similarity 39.3%; Pred. No. 4.5e-68;
Matches 286; Conservative 138; Mismatches 253; Indels 51; Gaps 13;

QY 8 RRDALTOLOV---RALMSAGSRAVA-----VDLGNKLEISS 43
DB 17 RRARFPAPLSPVGRAPFLSGAAAEVAQADAPPPPPGKRVLESFREFEIGGRVISFET 76
QY 44 GKLAEPDGSAAVVGSGTAVMT-AVSKTKPSQPMPLVVDYRQKAAAAGRIPTNYLR 102
DB 77 GKMAFPANGSVVISMDDTHVLSVAAKSSSEPRDFPLVVDYQEKQYAGVIPPYMR 136
QY 103 EVGSDKEILTSRIIDRSIRPLPAGYFYDTQV-----LCNLAVGVNPEVLAINGA 156
DB 137 EGAPKERELLCGRILIDRPIPLPPPGFYHEVQVNMATIIWNVSISSGKODPPVMAANAS 196
QY 157 SVSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNVVAGAPKSOIWMLEASA 216
DB 197 SAALMLSDIPWNGIIGVIRGRIDGNFVLANPTVDELGSLDLNVYA-CSRDKTLMDVQA 255
QY 217 ENLIQDFCHAIKVGKYYTOOIIQGIQOLVKETGVTKTPQKLFPPSPBEIVKTHKLAME 276
DB 256 REITERDLQGMKLAHAENAKCINPOLRLAKRAG-KKKKEKYKISLSDSKYKIRTLSEA 314
QY 277 RLVAFTDYHDKVSDEAVNKIRLDTBEOLKKEKFPADPYEIIIESFNVAKEVFRSIVL 336
DB 315 PIEEVFTDSTYGKEFGEALENITQSVAKLBECCDESLKFLHKAVDYTRKQVIRKRII 374
QY 337 NEYKRCGRDLTSLRNVSCVDMEFKTLHGSALFORGOTVLCVTPDSLESIGISD-QVI 395
DB 375 EKGIRVGRQLDREVRLPYCCSSSTYPIHGSALFSRGDTQVCTYTL-----GAGDQRL 429
QY 396 TAINGIKDKFMFLHYEPPPYATNEIGKVTGLNRRELHGALAEKALYVPVIR-RDPPTI 453
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:16:56 ; Search time 165 Seconds
(without alignment)
1532.752 Million cell updates/sec

Title: US-09-907-907a-42
Perfect score: 3557
Sequence: 1 DGFPLPRDRALTLQVRA.....TAVLLHNTQLDNERLITLLP 705

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3557	100.0	705	3 AAB08732	AAB08732 Amino aci
2	3556	100.0	722	7 ADC39094	ADC39094 Novel hum
3	3498	98.3	783	5 ABP69432	ABP69432 Human pol
4	3030	85.2	675	5 ABG30875	ABG30875 Human pol
5	2289	64.4	540	7 ADC39092	ADC39092 Novel hum
6	2285	64.2	504	4 AAB92684	AAB92684 Human pro
7	2285	64.2	504	7 ADJ69985	ADJ69985 Human hea
8	1982	55.7	899	4 ABG17275	ABG17275 Novel hum
9	1982	55.7	899	4 ABG08546	ABG08546 Novel hum
10	1982	55.7	899	7 ADE08997	ADE08997 Novel pro
11	1964	55.2	476	7 ADE07933	ADE07933 Novel pro
12	1917	53.9	439	4 ABG17276	ABG17276 Novel hum
13	1842.5	51.8	748	4 ABB58546	ABB58546 Drosophi
14	1238.5	34.5	541	4 ABG08547	ABG08547 Novel hum
15	1204.5	33.9	703	6 ABU24041	ABU24041 Protein e
16	1181.5	33.5	696	4 AAU33947	AAU33947 Staphyloc
17	1181.5	33.5	698	4 AAU36732	AAU36732 Staphyloc
18	1190.5	33.5	698	6 AAB16489	AAB16489 Protein e
19	1190.5	33.5	698	6 ABM73120	ABM73120 Staphyloc
20	1190.5	33.5	698	7 ADD26218	ADD26218 Staphyloc
21	1190.5	33.5	698	7 ADD26220	ADD26220 Staphyloc
22	1165.5	32.8	701	6 ABU43024	ABU43024 Protein e
23	1165.5	32.7	706	6 ABP40063	ABP40063 Staphyloc
24	1161.5	32.6	711	6 ABU43664	ABU43664 Protein e
25	1161	32.6	711	6 ABU47605	ABU47605 Protein e

26	1158	32.6	721	4 AAU38175	AAU38175 Salmonell
27	1157	32.5	729	6 ABU33310	ABU33310 Protein e
28	1156	32.5	1034	2 AAU03792	AAU03792 S. aureus
29	1155	32.5	703	6 ABU25139	ABU25139 Protein e
30	1155	32.5	734	4 AAU34719	AAU34719 E. coli c
31	1155	32.5	734	6 ABU28778	ABU28778 Protein e
32	1150.5	32.3	719	6 ABO67122	ABO67122 Klebsiell
33	1147	32.2	711	6 ABU31568	ABU31568 Protein e
34	1146.5	32.2	713	6 ABU19845	ABU19845 Protein e
35	1146	32.2	705	6 ABU50011	ABU50011 Protein e
36	1140	32.0	709	6 ABU40849	ABU40849 Protein e
37	1140	32.0	720	7 ADF05453	ADF05453 Bacteri
38	1135.5	31.9	719	6 ABU23367	ABU23367 Protein e
39	1135	31.9	712	6 ABU28271	ABU28271 Protein e
40	1134	31.9	709	4 AAU35396	AAU35396 Haemophil
41	1134	31.9	709	6 ABU30185	ABU30185 Protein e
42	1131.5	31.8	771	6 ABU20913	ABU20913 Protein e
43	1129.5	31.8	708	7 ADH88109	ADH88109 Enterococ
44	1128.5	31.7	704	4 AAU35289	AAU35289 Enterococ
45	1128.5	31.7	704	6 ABU14497	ABU14497 Protein e

ALIGNMENTS

RESULT 1	AAU38175
ID	AAU38175 standard; protein: 705 AA.
XX	AAU38175:
AC	02-JAN-2001 (first entry)
DT	
XX	
DB	Amino acid sequence of a human OLD-35 polypeptide.
XX	
KW	OLD-35; OLD-64; OLD-137; OLD-139; OLD-142; OLD-175; cancerous phenotype;
KW	cellular senescence; terminal differentiation; growth suppression;
KW	aging process; type I interferon; cancer cell; tissue regeneration; ss.
XX	
OS	Homo sapiens.
PN	MO200046391-A2.
PD	
XX	10-AUG-2000.
XX	
PP	02-FEB-2000; 2000MO-US002920.
XX	
PR	02-FEB-1999; 99US-00243277.
XX	
PA	(UYCO) UNIV COLUMBIA NEW YORK.
XX	
FI	Fleher PB, Leszczynska M;
DR	WPI; 2000-532905/48.
XX	N-PSDB; AAA64608.
PT	Novel isolated nucleic acid encoding an OLD-35 or OLD-64 protein useful
PT	in the treatment and detection of e.g. cancer and diseases involving
XX	cellular senescence.
PS	Disclosure; Fig 9B; 115pp; English.
XX	
CC	The specification describes OLD-35, OLD-64, OLD-137, OLD-139, OLD-142 and
CC	OLD-175 proteins. The OLD nucleic acids are useful for reversing the
CC	cancerous phenotype of a cancer cell, determining if a cell is senescent,
CC	growth arrested or terminally differentiated. They are also useful for
CC	reversing the aging process in a cell and degrading specific RNAs in a
CC	cell. The genes may also be used as a diagnostic indicator of cellular
CC	senescence, terminal differentiation and/or growth suppression and as a
CC	marker to identify drugs or small molecules that will induce or inhibit
CC	cellular senescence or terminal differentiation and type I interferons.
CC	The combination of Old-35 with other interacting proteins is useful for
CC	targeting the differentiation of specific cells. Old-35 can be used to

CC selectively stabilize specific mRNAs containing adenoviral rich 3' UTRs.
 CC The Old proteins are useful for reversing the cancerous phenotype of a
 CC cancer cell and inhibiting the growth of a cancer cell. They are also
 CC useful for regenerating tissue. The present sequence represents an Old-35
 CC polypeptide

XX Sequence 705 AA:

Query Match 100.0%; Score 3557; DB 3; Length 705;

Best Local Similarity 100.0%; Pred. No. 1,7e-301; Mismatches 0; Indels 0; Gaps 0;

Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPFLLPRDRALTOLOVRALMSSAGSRAVAVDIGNRKLEISGSKLARFADGSAVVOGSD 60
 DB 1 DGPFLLPRDRALTOLOVRALMSSAGSRAVAVDIGNRKLEISGSKLARFADGSAVVOGSD 60
 QY 61 TAVWTVASKTSPSPQMPVLVDYRQAAAGRIPTNYLRKRGVTSDEKELTSTRIDRS 120
 DB 61 TAVWTVASKTSPSPQMPVLVDYRQAAAGRIPTNYLRKRGVTSDEKELTSTRIDRS 120
 QY 121 IRPLFPAGFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
 DB 121 IRPLFPAGFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
 QY 181 GEYVNPFTKREMSSTNLVVAAPKSOIYMLEASAENILQDPFGHAIKVGKTTQOIIO 240
 DB 181 GEYVNPFTKREMSSTNLVVAAPKSOIYMLEASAENILQDPFGHAIKVGKTTQOIIO 240
 QY 241 GIQOLVETGVTKTPOKLTFTSPSEIVYTHKLMERYAFETQYEHDKVSRDAVVKIR 300
 DB 241 GIQOLVETGVTKTPOKLTFTSPSEIVYTHKLMERYAFETQYEHDKVSRDAVVKIR 300
 QY 301 LDTEEOQKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCDSRDLTSLNVSCEVDMF 360
 DB 301 LDTEEOQKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCDSRDLTSLNVSCEVDMF 360
 QY 361 KTLHGSALEFORQOVLCCTVTPDSLESGIKSDQVITYAINGIKDNFMHYEPFYATNEI 420
 DB 361 KTLHGSALEFORQOVLCCTVTPDSLESGIKSDQVITYAINGIKDNFMHYEPFYATNEI 420
 QY 421 GKTVGLNRRELGHGALAEKALYPIPRDPERTIRVTSFVLESNSSMAACGGSILAMD 480
 DB 421 GKTVGLNRRELGHGALAEKALYPIPRDPERTIRVTSFVLESNSSMAACGGSILAMD 480
 QY 481 SGVPISSAVAGVALGLVTKTDPKGEIIEYRLTLGIEEDYNDMDPKIAGTKKGIITAL 540
 DB 481 SGVPISSAVAGVALGLVTKTDPKGEIIEYRLTLGIEEDYNDMDPKIAGTKKGIITAL 540
 QY 541 QADIKLPGIPIKIYWEAIOQASVAKKEILOIMNTTISKPRASRENGPVETVOPLSKR 600
 DB 541 QADIKLPGIPIKIYWEAIOQASVAKKEILOIMNTTISKPRASRENGPVETVOPLSKR 600
 QY 601 AKFVPGSGYNKKIOAETGVVISOVDETFSEVFPPTBSVMEHARDPTEICKDQEOOLE 660
 DB 601 AKFVPGSGYNKKIOAETGVVISOVDETFSEVFPPTBSVMEHARDPTEICKDQEOOLE 660
 QY 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLANTQLDNERLNIILP 705
 DB 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLANTQLDNERLNIILP 705

RESULT 2
 ADC39094
 ID ADC39094 standard; protein; 722 AA.

XX AC ADC39094;
 XX DT 18-DEC-2003 (first entry)
 XX DE Novel human NOVX polypeptide SEQ ID NO: 36.
 XX KW antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
 KW nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;

KW antiaesthetic; antiinflammatory; hypotensive; antiarteriosclerotic;
 KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
 KW lymphoma; urogen cancer; prostate cancer; dyslipidemia; anorexia;
 KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
 KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
 KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
 KW Albright hereditary osteodystrophy.

XX Homo sapiens.

OS W02003010327-A2.

XX 06-FEB-2003.

XX 02-MAY-2002; 2002WO-US014199.

XX 02-MAY-2001; 2001US-0288063P.

XX 03-MAY-2001; 2001US-0288395P.

XX 07-MAY-2001; 2001US-0289087P.

XX 09-MAY-2001; 2001US-0289817P.

XX 09-MAY-2001; 2001US-0289818P.

XX 11-MAY-2001; 2001US-0290194P.

XX 14-MAY-2001; 2001US-0290753P.

XX 15-MAY-2001; 2001US-0291181P.

XX 16-MAY-2001; 2001US-0291243P.

XX 18-MAY-2001; 2001US-0292001P.

XX 21-MAY-2001; 2001US-0292374P.

XX 22-MAY-2001; 2001US-0292587P.

XX 23-MAY-2001; 2001US-0293107P.

XX 25-MAY-2001; 2001US-0293747P.

XX 29-MAY-2001; 2001US-0294109P.

XX 29-MAY-2001; 2001US-0294110P.

XX 30-MAY-2001; 2001US-0294434P.

XX 31-MAY-2001; 2001US-0294827P.

XX 12-JUL-2001; 2001US-0304879P.

XX 31-JUL-2001; 2001US-0308901P.

XX 14-AUG-2001; 2001US-0312270P.

XX 17-AUG-2001; 2001US-0313416P.

XX 10-SEP-2001; 2001US-0318463P.

XX 27-SEP-2001; 2001US-0325683P.

XX 18-OCT-2001; 2001US-0330292P.

XX 28-NOV-2001; 2001US-0333873P.

XX 03-DEC-2001; 2001US-0336909P.

XX 03-DEC-2001; 2001US-0337552P.

XX 21-FEB-2002; 2002US-0359245P.

XX 01-MAY-2002; 2002US-00136826.

(CUBA-) CURAGEN CORP.

XX Miller CE, Kekuda R, Malysankar UM, Li L, Pena CE, Spytek KA;

PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zeinusen BD;

PI Patunrajan M, Anderson DW, Mezes PS, Peyman JA, MacDougall JR;

PI Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;

PI Edinger SR, Ellerman K;

XX WPI; 2003-239445/23.

XX N-PSDB; ADC39093.

XX New NOVX polypeptides and polynucleotides, useful in gene therapy,

XX particularly for treating or preventing a syndrome associated with a

XX human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,

XX hypertension or hemophilia.

XX Claim 1; SEQ ID NO 36; 748bp; English.

XX The invention relates to new isolated NOVX polypeptides, the genes

XX encoding them or sequences having at least 95% identity to the amino acid

XX or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,


```
QY 181 GEYVNPFRKEMSSSTLNLVVAGAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQOIIO 240
D 198 GEYVNPFRKEMSSSTLNLVVAGAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQOIIO 257
QY 241 GICQVNVETGVTKTPOKLTFTSPSEIVKYTHKLAMERLYANFTPYEDHKVSRDAVVKIR 300
D 258 CIOQLVETGVTKTPOKLTFTSPSEIVKYTHKLAMERLYANFTPYEDHKVSRDAVVKIR 317
QY 301 LDTEQKKEKPEADPEYIEIESFNVAKEVERSVIVNEEYKRCGRDLTSLRNSCEVDMF 360
D 318 LDTEQKKEKPEADPEYIEIESFNVAKEVERSVIVNEEYKRCGRDLTSLRNSCEVDMF 377
QY 361 KTLHGSAIPORGQOVLCTVTFDSLBSGKSDQVITAINGIKDNFPMHYEPYATNEI 420
D 378 KTLHGSAIPORGQOVLCTVTFDSLBSGKSDQVITAINGIKDNFPMHYEPYATNEI 437
QY 421 GKATGRLRRELGHGALAEKALYPIPRDPFTIRVTSVLESNSSSMASACGSLALMD 480
D 438 GKATGRLRRELGHGALAEKALYPIPRDPFTIRVTSVLESNSSSMASACGSLALMD 497
QY 481 SGVFISSAVAGVAGLVTKTDPKEGEIEDYRLTDILGIEDYNGDMDFKIGTNKGITL 540
D 498 SGVFISSAVAGVAGLVTKTDPKEGEIEDYRLTDILGIEDYNGDMDFKIGTNKGITL 557
QY 541 QADIKLPKIPKIYMEALIQASVAKKEILOIMNKTISKPRASRKEGPPVETVQVPLSKR 600
D 558 QADIKLPKIPKIYMEALIQASVAKKEILOIMNKTISKPRASRKEGPPVETVQVPLSKR 617
QY 601 AKFPGPGYNLKKIOAETGVTSQVDETFSPVAPTPSVNHEARDFTTEICKDOEOOLE 660
D 618 AKFPGPGYNLKKIOAETGVTSQVDETFSPVAPTPSVNHEARDFTTEICKDOEOOLE 677
QY 661 FGAVYTTATTEIRDTGVNVLKYPNMTAVLLHNTOLDNERL 700
D 678 FGAVYTTATTEIRDTGVNVLKYPNMTAVLLHNTOLDNERL 717

RESULT 4
ABG30875
ID ABG30875 standard; protein; 675 AA.
XX
AC ABG30875;
XX
D 21-OCT-2002 (first entry)
XX
D Human polynucleotide phosphorylase 74.25.
XX
KW Human; enzyme; polynucleotide phosphorylase 74.25; malignant tumour;
KW haemopathy; human immunodeficiency virus infection; HIV;
KW immunological disease; inflammation.
XX
OS Homo sapiens.
XX
FN CN1341720-A.
PD 27-MAR-2002.
XX
PF 05-SEP-2000; 2000CN-00119892.
XX
PR 05-SEP-2000; 2000CN-00119892.
XX
(SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
PI Mao Y, Xie Y;
XX
XX WPI, 2002-501203/54.
DR N-PSDB; ABK89228.
XX
XX New polypeptide-polynucleotide phosphorylase 74.25 for treating malignant
PT tumor, hemopathy, human immunodeficiency virus infection, immunological
PT disease and various inflammations.
XX
```

```
PS Claim 1; Page 25-26 (disclosure); 33pp; Chinese.
XX
CC The present invention discloses a new polypeptide-polynucleotide
CC phosphorylase 74.25, a polynucleotide encoding the polypeptide and a
CC method for producing the polypeptide using DNA recombination technology.
CC The invention also discloses a method for curing several diseases, such
CC as malignant tumour, haemopathy, human immunodeficiency virus (HIV)
CC infection, immunological disease and various inflammations by using the
CC polypeptide. The invention also discloses an antagonist for resisting the
CC polypeptide and its therapeutic action, and also discloses the
CC application of the polynucleotide encoding the new polynucleotide
CC phosphorylase 74.25. The present sequence represents the polynucleotide
CC phosphorylase 74.25
XX
SQ Sequence 675 AA;
XX
Query Match 85.2%; Score 3030; DB 5; Length 675;
Best Local Similarity 94.5%; Pred. No. 2e-255;
Matches 602; Conservative 3; Mismatches 4; Indels 28; Gaps 1;
QY 64 MTAVSRTKPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVSTSDKEILTSRIIDRSIRP 123
D 1 MTAVSRTKPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVSTSDKEILTSRIIDRSIRP 60
QY 124 LFPAGYPTDPOVLCNLLAVDGNBPDVLAINGASVALSLSDIPMNGVGAVRIGIIDGEY 183
D 61 LFPAGYPTDPOVLCNLLAVDGNBPDVLAINGASVALSLSDIPMNGVGAVRIGIIDGEY 120
QY 184 VNPFRKEMSSSTLNLVVAGAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQOIIO 243
D 121 VNPFRKEMSSSTLNLVVAGAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQOIIO 180
QY 244 QLVKETGVTRTPOKLTFTSPSEIVKYTHKLAMERLYANFTPYEDHKVSRDAVVKIRLDT 303
D 181 QLVKETGVTRTPOKLTFTSPSEIVKYTHKLAMERLYANFTPYEDHKVSRDAVVKIRLDT 240
QY 304 EBQKKEKPEADPEYIEIESFNVAKEVERSVIVNEEYKRCGRDLTSLRNSCEVDMFKTL 363
D 241 EBQKKEKPEADPEYIEIESFNVAKEVERSVIVNEEYKRCGRDLTSLRNSCEVDMFKTL 300
QY 364 HGSALFORGQOVLCTVTFDSLBSGKSDQVITAINGIKDNFPMHYEPYATNEIGKV 423
D 301 HGSALFORGQOVLCTVTFDSLBSGKSDQVITAINGIKDNFPMHYEPYATNEIGKV 360
QY 424 TGLNRRELGHGALAEKALYPIPRDPFTIRVTSVLESNSSSMASACGSLALMDSGV 483
D 361 TGLNRRELGHGALAEKALYPIPRDPFTIRVTSVLESNSSSMASACGSLALMDSGV 420
QY 484 PISAVAGVAGLVTKTDPKEGEIEDYRLTDILGIEDYNGDMDFKIGTNKGITLQAD 543
D 421 PISAVAGVAGLVTKTDPKEGEIEDYRLTDILGIEDYNGDMDFKIGTNKGITLQAD 480
QY 544 IKLPKIPKIYMEALIQASVAKKEILOIMNKTISKPRASRKEGPPVETVQVPLSKRAKF 603
D 481 IKLPKIPKIYMEALIQASVAKKEILOIMNKTISKPRASRKEGPPVETVQVPLSKRAKF 527
QY 604 VPGGYNLKKIOAETGVTSQVDETFSPVAPTPSVNHEARDFTTEICKDOEOOLEFGA 663
D 528 -----GVTSQVDETFSPVAPTPSVNHEARDFTTEICKDOEOOLEFGA 572
QY 664 VYTATTEIRDTGVNVLKYPNMTAVLLHNTOLDNERL 700
D 573 VYTATTEIRDTGVNVLKYPNMTAVLLHNTOLDNERL 609

RESULT 5
ADC39092
ID ADC39092 standard; protein; 540 AA.
XX
AC ADC39092;
XX
D 18-DEC-2003 (first entry)
XX
```

DE Novel human NOVX polypeptide SEQ ID NO: 34.
 XX anti-diabetic; cytoprotective; immunomodulator; anorectic; antilipemic;
 KW neurotrophic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
 KW antineoplastic; antinflammatory; hypotensive; antiarteriosclerotic;
 KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
 KW lymphoma; uterine cancer; prostate cancer; dyslipidemia; anorexia;
 KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
 KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
 KW hyperextension; arteriosclerosis; hemophilia; graft-versus-host disease;
 KW Alldright hereditary osteodystrophy.
 XX Homo sapiens.
 XX WO2003010327-A2.
 PD 06-FEB-2003.
 XX 02-MAY-2002; 2002WO-US014199.
 PF 02-MAY-2001; 2001US-0288063P.
 PR 03-MAY-2001; 2001US-0288395P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 09-MAY-2001; 2001US-0289817P.
 PR 09-MAY-2001; 2001US-0289818P.
 PR 11-MAY-2001; 2001US-0290194P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291181P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 25-MAY-2001; 2001US-0293747P.
 PR 29-MAY-2001; 2001US-0294109P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294343P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 12-JUL-2001; 2001US-0304879P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 14-AUG-2001; 2001US-0312270P.
 PR 17-AUG-2001; 2001US-0313416P.
 PR 10-SEP-2001; 2001US-0318463P.
 PR 27-SEP-2001; 2001US-0325683P.
 PR 18-OCT-2001; 2001US-0330293P.
 PR 28-NOV-2001; 2001US-0333873P.
 PR 03-DEC-2001; 2001US-0336909P.
 PR 03-DEC-2001; 2001US-0337552P.
 PR 21-FEB-2002; 2002US-0359245P.
 PR 01-MAY-2002; 2002US-00136826.
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 PI Muller CE, Kekuda R, Malyankar UM, Li L, Pena CE, Spytek KA;
 PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zechsen BD;
 PI Paturajan M, Anderson DW, Mezes PS, Peyman JA, MacDougall JR;
 PI Padigaru M, Raseelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
 PI Edinger SR, Ellerman K;
 XX
 XX WPI; 2003-239445/23.
 DR N-PSDB; ADC39091.
 XX
 XX New NOVX polypeptides and polynucleotides, useful in gene therapy,
 PT particularly for treating or preventing a syndrome associated with a
 PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
 PT hyperextension or hemophilia.
 XX
 XX Claim 1; SEQ ID NO 34; 748bp; English.
 CC The invention relates to new isolated NOVX polypeptides, the genes
 CC encoding them or sequences having at least 95% identity to the amino acid
 CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
 CC particularly in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
 CC treating, preventing or alleviating pathology associated with NOVX
 CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
 CC polypeptide are especially useful for treating or preventing e.g.
 CC diabetes, obesity, cancers (e.g. lymphoma, uterine cancer or prostate
 CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
 CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
 CC disease, multiple sclerosis, hypertension, arteriosclerosis, hemophilia,
 CC graft-versus-host disease or Alldright hereditary osteodystrophy. The DNA
 CC encoding the protein is useful in gene therapy for treating the above
 CC conditions. These are also useful in developing powerful assay system for
 CC functional analysis of various human disorders, as well as in diagnostic
 CC applications. This sequence represents one of the NOVX proteins of the
 CC invention.
 CC
 XX Sequence 540 AA;
 XX
 XX Query Match 64.4%; Score 2289; DB 7; Length 540;
 XX Best Local Similarity 86.8%; Pred. No. 9e-191;
 XX Matches 455; Conservative 22; Mismatches 47; Indels 0; Gaps 0;
 QY 1 DGPFLPRRRALTOLOVRALMSAGSRAVAVDIGNRLKLSGKLARPADGSAVVGSGD 60
 DB 17 DGPECLGMDRALTOLOVRALLSTSGSAVAMDIGNKRLKLSGKLARFANGSAVVGSGD 76
 QY 61 TAVWTVASKTPSPSPQFMPVVDYRQKAAAGRIPTNYLAREGTSKDELTRIDRS 120
 DB 77 TAVMATASKTPSPSPQFMPVVDYRQKAAAGRIPTNYLAREGTSKDELTRIGVDCS 136
 QY 121 IRPLFPAGYFYDTQVLNLAVDGNEPDVLAINGASVALSLSDIPNNGPVGAVRIGIID 180
 DB 137 VRIFPAGYFCDIVLNLAVDGINELDVLAINGNSVALSLSDILNNGPVGTRIGMTD 196
 QY 181 GEVYNPFRKEMSSSTNLVVAAGPKSQIYMLKASANILOQDPCHAIKGVKTYQOIIQ 240
 DB 197 GECVNNPFRKEMSSSTNLVVAAGPKSQIYMLKASANILOQDPCHAIKGVKTYQOIIQ 256
 QY 241 GIQQLVKEIGVTKTPKLPSPSEIYKTHKLMERLYAVFTYEHDKVSRDEAVNKIR 300
 DB 257 GIQQLVKEIGVTKTPKLPSPSEIYKTHKLMERLYAVFTYEHDKVSRDEAVNKIR 316
 QY 301 LDTEQLKEKEPEADPYEIIIESFNVAKEVFRSIVLNEYKRCDRDLTSLNVSCEVDMF 360
 DB 317 LDTEQLKEIPEBVDLYEIIIESFTVAKVFRSIIINEYKRCDRDLTSLNISCEDVMF 376
 QY 361 KTLHGSAIPORGQTVLCTTTPDSLESIGKSDQVITAINIKDKNFMILHFEPPYATNEI 420
 DB 377 KTLHGSELFGQGTQLCAVTFDLSSESIKIDRVITTINGIKDKNFMILHFEPPYATNEI 436
 QY 421 GKVTGLNRRELGHGALAEKALYPIPRDPPTTIVTSEVSESSGSSMAASCGGSLAMD 480
 DB 437 VKVTGINRRELGPDALEKALYPIPRDPPTTIVTSEVSESSGSSMAASCGGSLAMD 496
 QY 481 SGVPISSAVGVALGVLTKTPDKGKIEDYRLLDIIGIEDYNG 524
 DB 497 SGVPIKSAVGVAMGLATKTDLEKGEIEDYRLLDIIGIEDYNG 540
 XX
 XX RESULT 6
 XX AAB92684
 ID AAB92684 standard; protein: 504 AA.
 AC AAB92684;
 XX 26-JUN-2001 (first entry)
 DT
 XX Human protein sequence SEQ ID NO:11065.
 DE
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 KM Homo sapiens.
 XX

PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 PT
 PS
 XX Claim 8; SEQ ID NO 11065; 2537bp + Sequence listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH1628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 CC
 CC
 XX
 SQ Sequence 504 AA;
 Query Match 64.2%; Score 2285; DB 4; Length 504;
 Best Local Similarity 100.0%; Pred. No. 1.8e-190;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DGPFLPRDRALTQLQVRAIMSSAGSRAVAVDIGNRKLEISGKLAFAADGSAVVOGSD 60
 DB 18 DGPFLPRDRALTQLQVRAIMSSAGSRAVAVDIGNRKLEISGKLAFAADGSAVVOGSD 77
 QY 61 TAVVAVTSKTPSPSQMPPIVVDYRQKAAAGRIPTMYLIREVGTSGKELTSTRIDRS 120
 DB 78 TAVVAVTSKTPSPSQMPPIVVDYRQKAAAGRIPTMYLIREVGTSGKELTSTRIDRS 137
 QY 121 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVALINGASVALSLSDIPMNGPVGAVRIGIID 180
 DB 138 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVALINGASVALSLSDIPMNGPVGAVRIGIID 197
 QY 181 GEYVAVPTRKEMSSSTLNLVVAAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQIIQ 240
 DB 198 GEYVAVPTRKEMSSSTLNLVVAAPKSGIWMLEASAEENIIQQDFCHAIKVGKXTTQIIQ 257
 QY 241 GIGQLVAVETGVTKRTPOKLTTPSPBEIVKYTHKLAMEELVAVFTDYEHDKVSRDEAVNKIR 300

DB 258 GIGQLVAVETGVTKRTPOKLTTPSPBEIVKYTHKLAMEELVAVFTDYEHDKVSRDEAVNKIR 317
 QY 301 LDTEEQJKEKFPPEADPYEIIIESFNVAKEVFRSIVLNEYRCQGRDITSLRNSCEVDME 360
 DB 318 LDTEEQJKEKFPPEADPYEIIIESFNVAKEVFRSIVLNEYRCQGRDITSLRNSCEVDME 377
 QY 361 KTLHGSALFQRCGTQVCTYTPDSLESIGSDQVITINIKKNFMIAHEFPYATNEI 420
 DB 378 KTLHGSALFQRCGTQVCTYTPDSLESIGSDQVITINIKKNFMIAHEFPYATNEI 437
 QY 421 GKVTGLNRRRLGHALAEKALYPIPRDFP 450
 DB 438 GKVTGLNRRRLGHALAEKALYPIPRDFP 467
 RESULT 7
 ADJ69985
 ID ADJ69985 standard; protein; 504 AA.
 AC
 XX ADJ69985;
 AC
 DT 06-MAY-2004 (first entry)
 DT
 XX
 DE Human heat mitochondrial protein as a therapeutic target SeqId1791.
 XX
 KW mitochondrial; human; screening assay; diabetes mellitus;
 KW Huntington's disease; osteoarthritis;
 KW Leber's hereditary optic neuropathy; LHON;
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 KW neuroprotective; noctropic; antidiabetic; anticonvulsant; antiarthritic;
 KW osteopathic; ophthalmological; cytostatic.
 OS
 XX Homo sapiens.
 XX
 PN WO2003087768-A2.
 PD
 XX 23-OCT-2003.
 PD
 XX 04-APR-2003; 2003WO-US010870.
 PF
 XX 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-0389987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX
 PA (MITO-) MITOKOR.
 PA (BUCK-) BUCK INST AGE RES.
 PA
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DB;
 PI
 DR WPI; 2003-845369/78.
 DR
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 PT
 XX Claim 1; SEQ ID NO 1791; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, noctropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and

CC cytosolic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.

XX Sequence 504 AA:

Query Match 64.2%; Score 2285; DB 7; Length 504;

Best Local Similarity 100.0%; Pred. No. 1.8e-190;

Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 DGPFLRRRRALTOLOVRALMSAGRAVAVDIGNKLEISSGKLARFADGSAVVOGSD 60
DB 18 DGPFLRRRRALTOLOVRALMSAGRAVAVDIGNKLEISSGKLARFADGSAVVOGSD 77
QY 61 TAVVAVTSKTPSPQFMPVAVYVYRQKAAAGRIPTNYLAREVGTSDKEILTSRIIDRS 120
DB 78 TAVVAVTSKTPSPQFMPVAVYVYRQKAAAGRIPTNYLAREVGTSDKEILTSRIIDRS 137
QY 121 IRPLFPAGFYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
DB 138 IRPLFPAGFYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
QY 181 GEYVAVPTRKEMSSSTLNLVVAGAPKSOIWMLEASAEIILQODFCHAIKGVKTTQOIIO 240
DB 198 GEYVAVPTRKEMSSSTLNLVVAGAPKSOIWMLEASAEIILQODFCHAIKGVKTTQOIIO 257
QY 241 GIOOLVKEGTGKTPKOKFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVRDEAVNKIR 300
DB 258 GIOOLVKEGTGKTPKOKFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVRDEAVNKIR 317
QY 301 LDTEBOLKEKRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMF 360
DB 318 LDTEBOLKEKRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMF 377
QY 361 KTLHGSLFPQRCQOVCTTFTDLSGSIKSDQVITAINIKDKNFMLEHPPYATNEI 420
DB 378 KTLHGSLFPQRCQOVCTTFTDLSGSIKSDQVITAINIKDKNFMLEHPPYATNEI 437
QY 421 GKVTGLNRRELGHGALAEKALYPYIPRPDP 450
DB 438 GKVTGLNRRELGHGALAEKALYPYIPRPDP 467
```

RESULT 8

ABG17275 standard; protein; 899 AA.

XX ABG17275;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #17266.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Dermanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS81462.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20; SEQ ID NO 47634; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantifying a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 899 AA;

Query Match 55.7%; Score 1982; DB 4; Length 899;

Best Local Similarity 73.2%; Pred. No. 1.5e-163;

Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

```
QY 10 DRAITLOQVRALMSAGRAVAVDIGNKLEISSGKLARFADGSAVVOGSDTAVMTAVS 69
DB 440 DRAITLOQVRALMSAGRAVAVDIGNKLEISSGKLARFADGSAVVOGSDTAVMTAVS 499
QY 70 KTRPSQGFMPVAVYVYRQKAAAGRIPTNYLAREVGTSDKEILTSRIIDRSIRLFPAGY 129
DB 500 KTRPSQGFMPVAVYVYRQKAAAGRIPTNYLAREVGTSDKEILTSRIIDRSIRLFPAGY 521
QY 130 FYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEYVAVNPT 189
DB 522 -----VGMDGECVAVNPT 535
QY 190 KEMSSSTLNLVVAGAPKSOIWMLEASAEIILQODFCHAIKGVKTTQOIIOGIVKETS 249
DB 536 KEMSSSTLNLVVAGAPKSOIWMLEASAEIILQODFCHAIKGVKTTQOIIOGIVKETS 595
QY 250 GVTKRTPOKFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVRDEAVNKIRLDEBOLKE 309
DB 596 GVTKRTPOKFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVRDEAVNKIRLDEBOLKE 655
QY 310 KRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMFKTLHGSLF 369
DB 656 IRPEVDLYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMFKTLHGSEL 715
QY 370 QRGOTVLCVTPDPSLSGSIKSDQVITAINIKDKNFMLEHPPYATNEIGKVTGLNRR 429
DB 716 QRGOTVLCVTPDPSLSGSIKSDQVITAINIKDKNFMLEHPPYATNEIGKVTGLNRR 775
QY 430 ELHGALAEKALYPYIPRPDPFTIRVTSVLESGSSSMASACGSLALMDSGVPISAV 489
DB 776 ELHGALAEKALYPYIPRPDPFTIRVTSVLESGSSSMASAC----- 818
QY 490 AGVAVGLVTKDPEKGEIEDYRLLTDIAGIEDNGMDDFIAGTNKGIITLQODIKLPGI 549
DB 819 -GVAMGLATKTDLEKGEIEDYRLLTDIAGIEDNGMDDFIAGTNKGIITLQODIKLPGI 877
QY 550 PIKIWEAIOOASVAKK 566
:|||||:|||||:
```


DR WPI; 2003-569235/53

PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2541; 1177pp; English.

CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers o
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence was used in the
CC exemplification of the invention

SQ Sequence 899 AA;

Query Match	1982;	DB 7;	Length	899;			
Best Local Similarity	-73.2%;	Pred No	1.5e-163;				
Matches	408;	Conservative	29;	Indels	102;	Gaps	2
		Mismatches	29;				

QY	10	DRALITQLQYRALMSKSGSFAVVDLGNRLKLEISSGKLAFADQSAVQSDITVMTAVS	69
Db	440	DRALITQLQYRALMSKSGSFAVVDLGNRLKLEISSGKLAFADQSAVQSDITVMTAVS	459
QY	70	KTTPSPSPQMPLVVDVYRQKAAAGRLPTMYLRREVSTDEILTSRIIDRSINPLPAGY	129
Db	500	KTTPSPSPQMPLVVDVYRQKAAAGRLPTMYLRREVSTDEILTSRIIDRSINPLPAGY	521
QY	130	FYDTQVLNLNLAVDGVNEPDTLAINGASVALSLSDIPMNGPVGAVARIGITIDEEVYNPTR	169
Db	522	FYDTQVLNLNLAVDGVNEPDTLAINGASVALSLSDIPMNGPVGAVARIGITIDEEVYNPTR	535

Qy	190	LEMSSTLNLVYAGAPKSIQWLEBAENILQODPCHAIKVGVKTYQOIIIGIQQLVKET	249
Db	536	KEMSSSTLNVAAGAPKSIQWLEBAENILQODPCHAIKVGVKTYQOIIIGIQQLVKET	599
Qy	250	GVTKRTPOKLFTRPSBEIVYKTKHLAMERLYAAPTVDYEHDKVSNDKAVNKIRLDTBEOLKE	309
Db	596	GVTKRTPOKLFTRPSBEIVYKTKHLAMERLYAAPTVDYEHDKVSNDKAVNKIRLDTBEOLKE	655
Qy	310	KPEADPYEIIISFNVAKEVFRSITVLENYKRCDSGDLTSLRVSCSEVMDFKTLHGSAF	369
Db	656	IFPEVDLVETIIISFNVAKEVFRSITVLENYKRCDSGDLTSLRVSCSEVMDFKTLHGSELF	715

Qy	370	RGRCGRVLTCTYFSDLSSEGIKSDQVLTANGIKDKKPMILYEEPPATNIGVYTGILNR	429
		QNGQTQLLCALVFSDLSSESIKIDRVLTITNGIKDKKPMILYEEPPATINELVKTGTILNR	
Db	716		
Qy	430	ELGHGALAEKALYPVLPDPPTTIRTVSEVLSENGSSSMASACGSLALMDSGVPISSAV	489
		ELGPPALAEKALYPVLPDPPTTIRTVSEVLSENGSSSMASAC-----	
Db	776		818

Qy	490	AGVAGLGLTKTDPKGEKIEDYRLTIIIGIEDVNGSMDPKLAKTNGKITLADIDIKLPGI	549
Db	819	-GVAGKGLTKTDLPEKGEIEDYRLITIIIGLEAVNGSMDPKLAKTNGKITLADIDIKLPGI	877
Qy	550	PKITVMEALQQASVAKK	566
Db	878	TKKITVMEALQQASVAKR	894

RESULT 11
ADE07933
ID ADE07933 standard; protein; 476 AA.
XX
AC ADE07933;
XX
DT 29-JAN-2004 (first entry)
DE Novel protein (useful for identifying genetic disorders) #88

XX novel gene; novel protein; tissue marker; molecular weight marker
 KW chromosome marker; genetic disorder.
 XX
 OS Unidentified.
 XY

PN WO2003054152-A2

PD 03-JUL-2003

PF 10-DEC-2002

PR 10-DEC-2001; 2001US-0339739P.

PR	14-MAR-2002	2002US-0365091P
PR	14-MAR-2002	2002US-0365384P
PR	12-APR-2002	2002US-0372731P
PR	12-APR-2002	2002US-0372615P
PR	22-APR-2002	2002US-00128558
PR	24-APR-2002	2002US-0376045P

PA (HYSE-) HYSEQ INC
XX

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

DR WPI; 2003-569235/53
DR N-PSDB; ADE07022.

PT New polynucleotides, useful for expressing recombinant proteins for PT analysis, characterization or therapeutic use, or as markers for tissues PT in which the corresponding protein is preferentially expressed.

PS Claim 20; SEQ ID NO 999; 1177pp; English

CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence represents a protein
CC of the invention.

Sequence 476 AA,

Query Match	55.2%	Score 1664	DB 7	Length 476
Best Local Similarly	72.1%	Pred. No. 2e-162		
Matches 408; Conservative	16	Mismatches	30	Indels 112; Gaps 4

Oy DGEPLLPDRDRLALTOQLVALMSSAGSAVAVDLGNKRLTETISGGLARFADGASIVVSGD 60
| | | | |
Dd DGEPFLPRDRRLALTOQLVALMSSAGSAVAVDLGNRLP-----ALF-----FMPSGD 67
| | | | |
Oy TANMTAVASTKTSPSQPMFLPVVDYRQCKAAAGRLPTMYLARREVGTSDKEILTSRIIDRS 120
| | | | |
Dd TAAWTVASTKTSPSQPMFLPVVDYRQCAA----- 98
| | | | |

Qy	12.	IRRLPAGFYFDYDTQVLCNLMLADVGNBPDVLAINGASVALSLSDIPMNGPVGAVRISIID	160
Db	99	-----VGMTD	103
Qy	181	GEVYNPTKREKSSSTTLNVNAGARKSQIWNLEASAEMLIIODRECAIKYGVKTTQIIQ	240
Db	104	GECVNPTKREKSSSTTLNVNAGARKSQIWNLEASAEMLIIODRECAIKYGVKTTQIIQ	163

241 GIGQQLKEGNGYKRRFPQKLFPTSPBIVYCKTKTKLMEIRYAFPTDYEHDKXSRBAVVKIR 300
 Db 164 GIGQQLVKEIGVTKRTPQKLFPTSPBIVYCKTKTKLMEIRYAFPTDYEHDKXSRBAVVKIR 223
 Gy 301 LDNEQLKEKFPADPYEIIISFNVAKEVPSITLVNIEYKCDGRDITSLRNVSCEVDME 360

```
Db 224 LDTEBQKEKFEPEVDLYEIIIESFNTVAKEVFRSIIILNEYKRCGDRDLTLRNISCEVDMF 283
Qy 361 KTLHGSALFQRGQTVLCTVTFDSLESIGKSDQVITAINGIKDNKXNFMHLHYEPFYATNEI 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 284 KTLHGSSELPFGQGTQVLCVATFDSLESISKDRVITTTINGIKDNKXNFMHLHYEPFYATNEI 343
Qy 421 GKVTGLNRRELGHGALAEKALYPIVPRDPFTIRVTSEVLESNGSSSMASACGSLALMD 480
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 344 VKVTGIRNRRELGPALAEKALYPIVPRDPFIRVTSEVLESNGSSSMASAC----- 395
Qy 481 SGVPISSAVNAGVLTCTDPEKGEIEDYRLTDIIGIEDYNGMDMPKIGTKNGITLAL 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 396 -----GVANGLATKTDLEKGEIEDYRLTDIIGIEAYNGMDMPKIGTKNGITLAL 445
Qy 541 QADIKLPQIPKIKIYWEAIIQASVAKK 566
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 446 QADIKLPQITWKIKIYWEAIIQASVAKR 471

RESULT 12
ABG17276
ID ABG17276 standard; protein; 439 AA.
XX
AC ABG17276;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17267.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX N-PSDB; AAS81463.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 47635; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
```

```
CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 439 AA:
Query Match 53.9%; Score 1917; DB 4; Length 439;
Best Local Similarity 78.7%; Pred. No. 2.2e-158;
Matches 396; Conservative 6; Mismatches 15; Indels 86; Gaps 3;

Qy 1 DGPFLPRDRALTOLOVRALNWSAGRAVAVDGNRKLETSSGKLARFADGSAVNGSD 60
Db 18 DGPFLPRDRALTOLOVRALNWSAGRAVAVDGNRKLETSSGKLARFADGSAVNGSD 77
Qy 61 TAVVVTAVSKTKPSQSPQFPLVVDYRQKAAAGRIPTNYLRREVSTDKELTSRIDRS 120
Db 78 TTVVVTAVSKTKPSQSPQFPLVVDYRQKAA----- 108
Qy 121 IRPLFPAGYFYDQVLCNLLAVDGVNEPDLVAINGASVALSLSDIPMNGPVGAVRIGIID 180
Db 109 -----AVRIGIID 116
Qy 181 GEYVYNPTRKEMSSSTINLVVAGAPKQIYMLASAEINIIQDPFGHAKYGVYTOQIIQ 240
Db 117 GEYVYNPTRKEMSSSTINLVVAGAPKQIYMLEASAEINIIQDPFGHAKYGVYTOQIIQ 176
Qy 241 GIQQLVETGVTKRTPQKLFTPSPBEIVKYTHKLAMERLYAVFTDYEHDKYRDEAVVKIR 300
Db 177 GIQQLVETGVTKRTPQKLFTPSPBEIVKYTHKLAMERLYAVFTDYEHDKYRDEAVVKIR 236
Qy 301 LDTEBQKEKFPPEADPYEIIIESFNVAKEVFRSIVLNEYKRCGDRDITSLRNVSCEVDMF 360
Db 237 LDTEBQKEKFPPEADPYEIIIESFNVAKEVFRSIVLNEYKRCGDRDITSLRNVSCEVDMF 296
Qy 361 KTLHGSALFQRGQTVLCTVTFDSLESIGKSDQVITAINGIKDNKXNFMHLHYEPFYATNEI 420
Db 297 KTLHGSALFQRGQTVLCTVTFDSLESIGKSDQVITAINGIKDNKXNFMHLHYEPFYATNEI 356
Qy 421 GKVTGLNRRELGHGALAEKALYPIVPRDPFTIRVTSEVLESNGSSSMASACGSLALMD 480
Db 357 GKVTGLNRRELGHGALAEKALYPIVPRDPFTIRVTSEVLESNGSSSMASACRKPFI-- 414
Qy 481 SGVPISSAVA---GVALGLVTKT 500
Db 415 NGFRGSNFICCRPYKGLVTKT 437

RESULT 13
ABBS8546
ID ABBS8546 standard; protein; 748 AA.
XX
AC ABBS8546;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2430.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
```


Qy 450 PFTIRVTSEVLSSNGSSSMASACGSLALNDGVPISAVAGVAIGLVTKTDPKGEIED 509
 Db 419 PYTIRLVSEVLSSNGSTSQASVCGSTLALNDAGVPIKRPAGIAMGLITSEDLSEAV-- 476
 Qy 510 YRLITDLIGIEDYNGDMDFKIAGTKGITALQADIKLPGLPIKIWEAIOQASVAKKEIL 569
 Db 477 ---ITDIOGLEDFPGDMDFKVGAGTEKGITAIQVDTKIHGLSKYCICKTAINDARKARLFI 533
 Qy 570 QIMMKTSKPRASRKENGVPVETVQVPLSKRAKFFVPGGVNLKKLOAETGVTSIQVDEET 629
 Db 534 EKWACINERPKELSTAPRAYTINIDTDKIRTLIGTGKTTINKIEBTGVKIDIREDET 593
 Qy 630 FSVFAPTPSVHBADEFTIEICKDOEQLEFGAVYTATIEIRDGVWVLYPNMTAVL 689
 Db 594 VFVLSSDADSANRALKMIIDLTGD-----VKGGEVILGKVTKITNFGAFVEVLPGKEG-L 647
 Qy 690 LHNTQOLDNERLN 701
 Db 648 WHISKLDINKVN 659

Search completed: January 28, 2005, 19:31:24
 Job time : 170 secs

This Page Blank (Just to)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:18:21 ; Search time 44 Seconds

(without alignments)
1541.655 Million cell updates/sec

Title: US-09-907-907A-42

Sequence: 1 DGEPLPRRDRALTLQVRA.....TAVLHNTQLDNERLTLLP 705

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1365.5	38.4	991	2 T48631	polynucleotide pho
2	1204.5	33.9	703	2 B97123	polyribonucleotide
3	1190.5	33.5	698	2 B89901	polyribonucleotide
4	1166.5	32.8	713	2 A82586	polyribonucleotide
5	1166.5	32.8	713	2 F97368	polyribonucleotide
6	1161	32.6	711	2 AD0901	polynucleotide pho
7	1158.5	32.6	749	2 G97782	hypothetical prote
8	1158	32.6	734	2 H85979	polynucleotide pho
9	1158	32.6	734	2 E91134	polynucleotide pho
10	1156	32.5	810	2 G75320	polynucleotide pho
11	1153	32.4	718	2 S74509	polynucleotide
12	1152	32.4	734	2 H65106	polyribonucleotide
13	1150	32.3	714	2 AC3497	polyribonucleotide
14	1146	32.2	705	2 AC0424	polyribonucleotide
15	1145	32.2	704	2 G83950	polynucleotide pho
16	1140.5	32.1	745	2 B71654	polyribonucleotide
17	1140.5	32.1	775	2 E70320	polyribonucleotide
18	1138.5	32.0	718	2 AD2355	polyribonucleotide
19	1137	32.0	719	2 P82831	polynucleotide pho
20	1134	31.9	709	2 B64056	polyribonucleotide
21	1132	31.8	705	2 S70691	polyribonucleotide
22	1130	31.8	723	2 AG1603	polynucleotide pho
23	1127	31.7	723	2 AC1241	polynucleotide pho
24	1121	31.5	709	2 B82298	polyribonucleotide
25	1115	31.3	707	2 B84973	polyribonucleotide
26	1114.5	31.3	694	2 D86615	polyribonucleotide
27	1114.5	31.3	694	2 D72009	polyribonucleotide
28	1112	31.3	707	2 C81161	polyribonucleotide
29	1111	31.2	706	2 H81943	probable polyribon

30	1106	31.1	709	2 S38883	polyribonucleotide
31	1105	31.1	712	2 B87253	polyribonucleotide
32	1097	30.8	701	2 C83052	polyribonucleotide
33	1091	30.7	708	2 A72264	polynucleotide pho
34	1068.5	30.0	695	2 E71463	probable polyribon
35	1067.5	30.0	722	2 D70200	polyribonucleotide
36	1066	30.0	702	2 C71269	probable polyribon
37	1064	29.9	693	2 G81725	polyribonucleotide
38	1063.5	29.9	773	2 G86856	hypothetical prote
39	1054	29.6	737	2 D95068	polyribonucleotide
40	1053	29.6	775	2 D97936	polyribonucleotide
41	1029.5	28.9	719	2 G81332	polyribonucleotide
42	1016.5	28.6	897	2 T06540	polyribonucleotide
43	1005.5	28.3	739	2 T10932	polyribonucleotide
44	1000.5	28.1	777	2 T44900	probable polyribon
45	998.5	28.1	773	2 H87015	hypothetical prote

ALIGNMENTS

RESULT 1

T48631
polynucleotide phosphorylase - Arabidopsis thaliana
N/Alternate names: protein T15N1.70
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C/Accession: T48631
R/By: M. J. Murphy, G. J. Riley, P. J. Hudson, S. J. Bancroft, I. J. Mewes, H. W. J. Rudd, S. J. Le
submitted to the Protein Sequence Database, April 2000
A/Reference number: Z24493
A/Accession: T48631
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-991 <BEV>
A/Cross-references: UNIPROT: Q9S7G6; EMBL: AL163792
A/Experimental source: cultivar Columbia; BAC clone T15N1
C/Genetics:
A/Map position: 5
A/Intons: 107/3; 159/3; 216/3; 258/3; 312/3; 352/3; 398/3; 453/1; 495/1; 538/3; 612/1;
A/Note: T15N1.70
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match	38.4%	Score 1365.5	DB 2	Length 991
Best Local Similarity	41.7%	Pred. No. 5.4e-77		
Matches	287	Conservative 138	Mismatches 231	Indels 33
Gaps	11			
QY	24	SAGSRVAV---AVDLGNRLKLEISGKLARFADSGAVVQSGDTAVMTAVSKTRPSQF	78	
DB	44	SAGTKLIESFKBEFEVSGSRVVSFETGKLARFANGSVVLGMDETIVLSTVCAKTDSPDF	103	
QY	79	MLPVVDYRQKAAAGRIPTNYLREVGTSDEKILTSRIIDRSIRPLPAGFYTOVLGN	138	
DB	104	LPLTVDOEKQYAOGLIPNTYMRREGAPKKEELLCGRILIRPILPPTGTVHVOIMAS	163	
QY	139	LIAVGVNEPPVLAINGASVALISDI PMNGPVGAVIGIIGDGVVNPTRKEMSSSTLN	198	
DB	164	VLSGDKQDPITLAANSSAALMLSDVPGGIVIRIGICGFPVNPITMDELSSDLN	223	
QY	199	LVAAGPKSQIVMLASAEENITQDPCHAKV---GVKTOQIIIOGLOLVKETGVYKR	254	
DB	224	LIVA-CIRDKMTMDVQSRSEKDLAALRLARPAVKYLDPI---RLAEVAGQKK	278	
QY	255	TPQKLTFPSPFIVKYTKLAMEKLYAVFTYEHDKVSRDEAVNKIRIDTEQLKEKPEEA	314	
DB	279	-EYKLSMSDKTEKLVADLATRISVFTPSYKFKRGELDNIIGDVRVFEEDQD	337	
QY	315	DPEVITSPFNVAKEVRSIVLNEYKRCDSGDLTSLNVSCEVDMFTLGSALFQSGQT	374	
DB	338	SLSLPKAVDVRKKVRSKMSISDFVDSGHVDEVPFYCESHYLPALMSALFSKQDT	397	
QY	375	QVLCTVFDSLESGIKSDOVITAIINGIKDKNFMILHYEPYATNIEIGKVTGLNRELGHG	434	

Db 398 QVLTCTVTLGAPAEAKSLSLV----GPPKRFMLHYEPFPYCTNEBVGKGGANREVBHG 453
Qy 435 ALAEKALYVPIPRD--PFFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGV 492
Db 454 TLAKALLAVLPPEAPPYTIRINSEVWSSDGTSMSVCGSMAIMADAGIPLAHVAGV 513
Qy 493 AIGVLTDPKPGKEIEDYRLTLTDLIGIEDVNGMDFKIAGNKIGITALQADIKLPGPIK 552
Db 514 SVGLITVDPSGSEIKDYRIYTDILIGEDHIGDMDFKAGTRDVTVALQDLIKPAGIPLD 573
Qy 553 IYMEAIQOASVAKKEILQIMNKTSKPRASKENGVPVETVQVPLSKRAKVFVGXVNLK 612
Db 574 IVCSLENAARAKQIIDLHMERKNINSPGQGAVSPLATIKYSNDSLRPTIGPWGVLKR 633
Qy 613 KLAQETGVTLSQVDEEFPSPVAPTPSVHMEAR---DFITTEICKDDQEOLEFGAVYTATI 669
Db 634 KIEVETGARLS--LNGTLITIVAKNQDWEKAQEOQDFIL-----GRELVGVGYYKGTV 685
Qy 670 TEIRDGVVMKLYRPMTRAVLLHNTQLDNE 698
Db 686 SSIXEYGAFAVE--FPGQGGGLHMSLSHE 713

RESULT 2

B97123
polyribonucleotide nucleotidyltransferase [imported] - Clostridium acetobutylicum
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C/Accession: B97123
R;Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: B97123
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-703 <KUR>
A/Cross-references: UNIPROT:Q97145; GB:AE001437; PIDN:AAK79773.1; PID:g15024781; GSPDB:C
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC1808
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 33.9%; Score 1204.5; DB 2; Length 703;
Best Local Similarity 39.6%; Pred. No. 3.4e-67;
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

Qy 33 DLGRKLEISSGKLARFADGSAYVQSGDTAVMTAVSKTKPSPS-QFMPLVVDYRQKAA 91
Db 8 DIAGRKLKVECGKTGMLSNCAMFISYGGTVVMVNVNASEKREBGIDFPPLSIEYERQYS 67
Qy 92 AGRIPTYLVLRREVSTDSKEILTSRLIDRSIRLPFAGFYVTOYLCNLAAVGNBPVL 151
Db 68 VGKIPGFGVKEGPRSEKSIHARALDRPLRPKGRNDVQVCTVWSEQNLPIL 127
Qy 152 AINGASVALSLSDIPMNGPVAVRIGIIDGEYVNPTRKEMSSSTLNLVAGAPKSOIVM 211
Db 128 AMNGASVALSLSDIPPTTPVATVSGCIDGKFLVNPITLBEKSSLDLTVG-AINERYMM 186
Qy 212 LEASAEINLQODFCHAIKVGKYTQOIIQGIQQLVKTGVTKRPQKLPFSPPIVYKTH 271
Db 187 LEAQADEFEDLMTAALIDFGNACODIVAFQEKAMKEGKEKVPPE-LYHKEEIEKQVT 245
Qy 272 KLAHERLYAVFTVDYEHOKVSRDEAVNKRILDTESQLEKRPPEADPYEIISEFNVAKEVF 331
Db 246 EFAEESIKELM--YTTDERENLRLEIKESINSEPEKYPD--GCADDEVVYTLQKKV 302
Qy 332 RSIYLVNEKRCDCGRDLTSLRNVSCVDMFKTLHGSALFORQOTVLTCTVFPDSLESGIKS 391
Db 303 RNMILKHEHRDRDGRFPDEIRPISCDVDLLPRTHSGGLFTRGLTVQMYVTTL-----GPIGD 358
Qy 392 DQVITAINGIKDKNFMILHYEPFPYATNIEIGKVTGLNRRRELGHGALAEKALYVPIR--RDF 449

Db 359 AQVLDIGVSESKRYMHNYPFPYSTGEVRLKGPNNREIGHGALAKALVPLPSEBEF 418
Qy 450 PFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVIGVTTDPEKGEIED 509
Db 419 PYTIRLVSEVLSNGSTSOASVCGSSTLALMDAGPIRPAAGIAMGLTSEDSKEAV-- 476
Qy 510 YRLITDLIGIEDVNGMDFKIAGTNKIGITALQADIKLPGIPIKIYMEAIQOASVAKKEIL 569
Db 477 ---ITGLIEDFEGDMDFVAGTEKIGITAIQVDTKIHGLSKYCIKTAINDARKARLFIIL 533
Qy 570 QIMNKTSKPRASKENGVPVETVQVPLSKRAKVFVGXVNLKLAQETGVTLSQVDEET 629
Db 534 EKWACINPEPKELSTYAPRAYTINIDTKIRTLIGTGSNTINKITIEETGKVIDIREDT 593
Qy 630 FSVAPTPSVHMEARDFITTEICKDDQEOLEFGAVYTATIETIRDGVVMKLYRPMTRAVL 689
Db 594 VFVLTSSDADSANALAKMIDDLTGD----VKGEVYLGKTKTKTINPQAFVPLPGKSG-L 647
Qy 690 LHNTQLDNERLN 701
Db 648 VHSKLDINKVN 659

RESULT 3

B89901
polyribonucleotide nucleotidyltransferase [imported] - Staphylococcus aureus (strain N31)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C/Accession: B89901
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguci, ma, A.; Mitutani, O.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: B89901
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-698 <KUR>
A/Cross-references: UNIPROT:Q99U08; GB:BA000018; PID:g13701074; PIDN:BAB42369.1; GSPDB:G
A/Experimental source: strain N315
C/Genetics:
A/Gene: nupA
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 33.5%; Score 1190.5; DB 2; Length 698;
Best Local Similarity 39.1%; Pred. No. 2.5e-66;
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;

Qy 24 SAGRAVAVDLAGNRKLEISSGKLARFADGSAYVQSGDTAVMTAVSKTKPSPSQFMPLV 83
Db 2 SOEKVKTETMAGRSILTIETGQAKQANGAVLVRYGDTVALSTATASKEPRDGFPLTV 61
Qy 84 DYRQKAAAARIPNVIKREVGVSDKEILTSRLIIDSIRLPLFAGYTYDPQVLCNLAAVD 143
Db 62 NYEKMYAAKIPGFGFKRGGRGDDATLTARLIDRIRLPFKGXYGNDVQIMMYLSAD 121
Qy 144 GNVBPDLAINGASVALSLSDIPMNGPVAVRIGIIDGEYVNPTRKEMSSSTNLNVAG 203
Db 122 PDCSPQMAAIGSSMALSVSDIPPOGPIAGVNVGIDGKIINPTVEKSVSLDLEVAG 181
Qy 204 APSQIYMEASAEINLQODFCHAIKVG-----YKTYQOIIQGIQQLVKTGVTKRTP 256
Db 182 -HDAVVMVAVAGASEITEGEMLEAIFPGHEIGQLVDFQOQIYDHIQVKGQ----- 232
Qy 257 QKLPFS-----PEIVKTKHLAMERLYAVFTVDYEHNDVSRDEAVNKRILDTESQLE 309
Db 233 ---FIPEARDEALVERKSLTEBKGLKETVLT-----DKQORDEMDLNK---EELVNE 281
Qy 310 KFPKPADP-----YEIESNVVAKVEFRSIVLVNEKRCGRDLTSLRNVSCVDMFKTL 363
Db 282 FIDEDEPENLLIKEVYAILNELVKEVRRLIADKIRPDGRKRPDEIRPLDSEVGILPRT 341

```

Qy 364 HGSAALFORGQTVCTVTFDSLESIGKSDQVITAINIGIKDNFMLHYEPFPATNIEIKV 423
D 342 HGSGLFRGQTOALSVLTGLAL-----GDYQILDIGLPEBEKRFMHVNFPPSVGETPVP 397
Qy 424 TGLNRELIGHALAEKALYPIV--RDPFFITRVTSVLESNGSSSMASACGSLALMDS 481
D 398 RAPRRREIGHALAEKALYPIV--RDPFFITRVTSVLESNGSSSMASACGSLALMDS 457
Qy 482 GVPISAVAGVAGLVTGDEPKGRIEDYRLITDILGIEDNNGMDPFKAGNNGITATQ 541
D 458 GVPISAVAGVAGLVTGDEPKGRIEDYRLITDILGIEDNNGMDPFKAGNNGITATQ 510
Qy 542 ADIKLPGIPIKIWEALIQASVAKKEILQIMNKITISKRASRKENGPPVETVQVPLSKRA 601
D 511 MDIKIDGTRIELEALQARGLRINMHLQITDQRTLSAARPVMTIKEDKIR 570
Qy 602 KFPVGGVNLKKLQLETGVTISQVDETFVAPFPVMEHARDPITEICKDQEQULEF 661
D 571 DVIPGGKKIRIIEIDETGVKLDIEDQGTIFIGAVDQAMINRAEIEITR-----BAEV 625
Qy 662 GAVVTATITTEIRDTGVWVKYIPNMTAVLLHNTQDNEEL 700
D 626 GQYQATVTKRIEKYGAVGLPFGKDA-LHHSQISKNRI 663

```

RESULT 4

```

AE2586
polyribonucleotide nucleotidyltransferase pnpa [imported] - Agrobacterium tumefaciens (A
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AE2586
R/Mod: D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, P.; Giller, W.; Grant, C.; Guenther, D.; Kutayarin, T.; Levy, R.; Li, M.; McClie
; Karp, P.; Kometo, F.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreepan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577, MUID:21608550, PMID:11743153
A/Accession: AE2586
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-713 <KUR>
A/Cross-references: UNIPROT:Q8UJ56, GB:AB008668, PDB:1AAL1107.1, PDB:1G1738399, GSPDB:C
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A:Gene: pnpa
A:Map position: circular chromosome
C/Superfamily: polyribonucleotide nucleotidyltransferase

```

```

Query Match 32.8%; Score 1166.5; DB 2; Length 713;
Best Local Similarity 39.1%; Pred. No. 8.2e-65;
Matches 265; Conservative 133; Mismatches 251; Indels 29; Gaps 12;

Qy 29 AVAADLGRKLEISGKLARPADGSAVVOGSDTAVMTAVASKTSPSQ--FMPLVVDYRQ 87
D 6 SVEIEMAGRPKLETKGVAQADGAVATYGETVNLATVSAKSPKQGDFFPLTVNQE 65
Qy 88 KAAAGRIPTVYLRREVTSDEKITLSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGYNE 147
D 66 KTYAAGRIPGYFRREGRPSEKELTVSRIDRIRPLFPBEGYKNDTVVAVVIGHDLEND 125
Qy 148 PDVLAINGASVALSLDIPMNGPVGAVRIGIIDEGYVNPFRKEMSSSTLNLVYAGAPKS 207
D 126 PDVLSMVAASALTLGSIPEMGVPGAVGYNIEYVNLPHLDEMDSEVLDLVVAGT--QD 184
Qy 208 QIVWLEASENIILQODFCHAIKVGKVTQOIIQGIQQLVKEGTGVTKTPOKLFTPSPEIV 267
D 185 AVLVASEAKELEINMGAVMFGHGFQPIYDAI--IKLAEVAAKSPRE-FEPEDPSA 240
Qy 268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKLRLDTEBQLEKPEPA--DPYEIIIESFN 324

```

```

D 241 LENEMGLAET-ELRTAVKITEKARVAADAVTKTKVNAFLPEEGAKSPPEIGAVFK 299
Qy 325 VVAKEVPSIYLANEYKCGDGRDLTSRNVSCVDMFKTLHGSAALFORGQTVCTVTFDS 384
D 300 HLOAKIVANVLDTKSRIDGDLSTVAPRISEVGLIPRTHSALFTKGETAIYAVATIGT 359
Qy 385 LESGKSDQVITAINIGIKDNFMLHYEPFPATNIEIKGVTGLNRELIGHALAEKALYV 444
D 360 GE-----DEQVYDSLGMVKERFLHYNPPYPSVGETGMSGPRREIGHGLAMRAIRFM 415
Qy 445 IP--RDPFFITRVTSVLESNGSSSMASACGSLALMDSGVPISAVAGVAGLVTGDP 502
D 416 LPTAEPFPYTLRVVSELTESNGSSSMATVCGTSLALMDAGVPLKPVAGIMGIL----- 471
Qy 503 EKEGEDYRLITDILGIEDYNGMDPFKAGNNGITATQADIKLPGIPIKIWEALIQAS 562
D 472 ---EGERPAVSLDILGDEDLGMDPFKAGTADGITSIQMDIKIAGITEEMKALNEAQ 528
Qy 563 VAKKEILQIMNKITISKRASRKENGPPVETVQVPLSKRAKEVGGVNLKKLQLETGVTI 622
D 529 GGRKHILGEMANALTESGQGFAPRIEVMNIPVDKIRVIGSGKVTREIIVETGAKI 588
Qy 623 SQVDETFVAPFPVMEHARDPITEICKDQEQULEFQAVVTATITTEIRDTGVWVKY 682
D 589 NIEDGTIVKIASASGKEIEAKRWIHSIVAE-----EVGQIVGTVKTDAPGAFVNF 643
Qy 683 PNMTRAVLLHNTQDNEEL 700
D 644 GARDG-LVHISQLASERV 660

```

RESULT 5

```

P97368
polyribonucleotide nucleotidyltransferase (PA4740) [imported] - Agrobacterium tumefaciens
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: P97368
R/Goodner, B.; Hinkle, G.; Gatlung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2223-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359, MUID:21608551, PMID:11743194
A/Accession: P97368
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-713 <KUR>
A/Cross-references: UNIPROT:Q8UJ56, GB:AB007869, PDB:1AAK8903.1, PDB:1G15154950, GSPDB:C
A:Gene: AGR_C_124
C/Genetics:
A:Map position: circular chromosome
C/Superfamily: polyribonucleotide nucleotidyltransferase

```

```

Query Match 32.8%; Score 1166.5; DB 2; Length 713;
Best Local Similarity 39.1%; Pred. No. 8.2e-65;
Matches 265; Conservative 133; Mismatches 251; Indels 29; Gaps 12;

Qy 29 AVAADLGRKLEISGKLARPADGSAVVOGSDTAVMTAVASKTSPSQ--FMPLVVDYRQ 87
D 6 SVEIEMAGRPKLETKGVAQADGAVATYGETVNLATVSAKSPKQGDFFPLTVNQE 65
Qy 88 KAAAGRIPTVYLRREVTSDEKITLSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGYNE 147
D 66 KTYAAGRIPGYFRREGRPSEKELTVSRIDRIRPLFPBEGYKNDTVVAVVIGHDLEND 125
Qy 148 PDVLAINGASVALSLDIPMNGPVGAVRIGIIDEGYVNPFRKEMSSSTLNLVYAGAPKS 207
D 126 PDVLSMVAASALTLGSIPEMGVPGAVGYNIEYVNLPHLDEMDSEVLDLVVAGT--QD 184
Qy 208 QIVWLEASENIILQODFCHAIKVGKVTQOIIQGIQQLVKEGTGVTKTPOKLFTPSPEIV 267
D 185 AVLVASEAKELEINMGAVMFGHGFQPIYDAI--IKLAEVAAKSPRE-FEPEDPSA 240
Qy 268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKLRLDTEBQLEKPEPA--DPYEIIIESFN 324

```


Dh 241 LENEMGLAET-ELRTAYKITEKARAAVAVTKYKAHFLPEEGBAKSPSEIGAVFK 299
Qy 325 VVAKEVRSIYLANEKYKCDGDLTSLRNVSCVDMFTLHGSALFORGOTVCLTFVDS 384
Dh 300 HLOAKYIRANVLDKSIDGRDLSTVAPYSEVGLIRTRHOSALFTGTGEROAVVALLGT 359
Qy 385 LESGKSDOVITAIINGIKDKNFMLHYEPFPYATNEIGKVTGLNRELGHGALAEKALPV 444
Dh 360 GE-----DEQYVDLSLTGMVKERFLLHYNFPYPSVGETGMGSPGRREIGHGLAWRAIRPM 415
Qy 445 IP--RDPPFTRVTSVLENSGSSMASACGSLALMDSGVPISSAAGVAGIYVITDPE 502
Dh 416 LPTAEGQPPYTLRVAVSETESGSSSMATVCGTSLAMDAGVPLAKPVAIGIAMGLIL----- 471
Qy 503 EKEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITLALQADIKLPGIPIKIYMEALIOQAS 562
Dh 472 ----EGERFAYLSDILGDEBDHIGDMDFKAGADITSLQMDIKLAGITTEIMKIALDQAO 528
Qy 563 VAKKEIIQIMNKITSKPRASKENGPNVETVQVPLSKRAKFPVGGYNLKKLQAEVTYI 622
Dh 529 GGRGHILGEMANALITESRGQGEFAPRIEVMNIPVDKIREVIGSGKVIREEVETGAKI 588
Qy 623 SQVDEFFSPAPRPSVMEHARDITEICKDQEOLEFGAVYATITTEIRDTGVMKLY 682
Dh 589 NIBEDGTVKLASASGKEIEAKRKIHISIVAE-----EVGOIYEGTVKTADEFAGAFVNF 643
Qy 683 PNMTAVLLHNTQDNERL 700
Dh 644 GARDG-LVHISQIASERV 660

RESULT 6

AD0901
polynucleotide phosphorylase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium
C:/Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A/Note: This species has also been called *Salmonella typhi*
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Oct-2003
C/Accession: AD0901
R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Groh, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AD0901
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-711 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAD07802.1; PID:G16504350; GSPDB:GN00176
C/Genetics:
A/Gene: STY3463
C/Superfamily: polynucleotide nucleotidyltransferase

Query Match 32.6%; Score 1161; DB 2; Length 711;
Best Local Similarity 39.9%; Pred. No. 1.8e-64;
Matches 270; Conservative 118; Mismatches 249; Indels 40; Gaps 11;

Qy 35 GNRKLEISGKLAFADGASAVVSGDTAVMTAVSKTKRPSQ-FMPLVVDYROKAAAG 93
Dh 12 GQHTVLTETGMAAOATAAVVMSDDTAVFTVVGQKAKGQDFPLTVVYQERTYAAAG 71
Qy 94 RIFPNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGNPDTLAI 153
Dh 72 RIFSPFRREGRBEGEITLRLIDRPVPLFPBGFVNEVAVITVSVANQVNPDIYAM 131
Qy 154 NGASVALSLSDIPMNGPVGAVRIGITDGEVYVNTKREKSSSTLNLVAGAPKQIYMLE 213
Dh 132 IGASVALSLSDIPMNGPVGAVRIGITDGEVYVNTKREKSSSTLNLVAGAPKQIYMLE 190
Qy 214 ASANIIQOQPCFCAIKYGVKTTQOIIQOIVVETGVTKTKTPQKLTPSPSEIYK----- 268

Dh 191 SEAEILSEDTMLCAVVPFGHEQOQVVIQALINDLYKEACKPRWDQ-----PEAVNDALNA 244
Qy 269 YTKLAMERL---YAVFTDYEDHKVSRDEAVNKIRLDTESQLEKEPEADPYEIISEFNV 325
Dh 245 RVAALESRLSDAYRI-----TDKQERYAQVDVAKSETIRQLIAEDDTLANEGEILHA 299
Qy 326 VAKKEVRSIYLANEKYKCDGDLTSLRNVSCVDMFTLHGSALFORGOTVCLTFVDS 385
Dh 300 HLOAKYIRANVLDKSIDGRDLSTVAPYSEVGLIRTRHOSALFTGTGEROAVVALLGT 359
Qy 385 LESGKSDOVITAIINGIKDKNFMLHYEPFPYATNEIGKVTGLNRELGHGALAEKALPV 444
Dh 360 GE-----DEQYVDLSLTGMVKERFLLHYNFPYPSVGETGMGSPGRREIGHGLAWRAIRPM 415
Qy 445 IP--RDPPFTRVTSVLENSGSSMASACGSLALMDSGVPISSAAGVAGIYVITDPE 502
Dh 416 LPTAEGQPPYTLRVAVSETESGSSSMATVCGTSLAMDAGVPLAKPVAIGIAMGLIL----- 471
Qy 503 EKEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITLALQADIKLPGIPIKIYMEALIOQAS 562
Dh 472 ----EGERFAYLSDILGDEBDHIGDMDFKAGADITSLQMDIKLAGITTEIMKIALDQAO 528
Qy 563 VAKKEIIQIMNKITSKPRASKENGPNVETVQVPLSKRAKFPVGGYNLKKLQAEVTYI 622
Dh 529 GGRGHILGEMANALITESRGQGEFAPRIEVMNIPVDKIREVIGSGKVIREEVETGAKI 588
Qy 623 SQVDEFFSPAPRPSVMEHARDITEICKDQEOLEFGAVYATITTEIRDTGVMKLY 682
Dh 589 NIBEDGTVKLASASGKEIEAKRKIHISIVAE-----EVGOIYEGTVKTADEFAGAFVNF 643
Qy 683 PNMTAVLLHNTQDNERL 700
Dh 644 GARDG-LVHISQIASERV 660

RESULT 7

G97782
hypothetical protein pnp [imported] - *Rickettsia conorii* (strain Malish 7)
C:/Species: *Rickettsia conorii*
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: G97782
R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: G97782
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-749 <KOR>
A/Cross-references: UNIPROT:Q92HV7; GB:AE006914; PIDN:AAU03201.1; PID:G15619751; GSPDB:G A/Gene: pnp
C/Superfamily: polynucleotide nucleotidyltransferase

Query Match 32.6%; Score 1158.5; DB 2; Length 749;
Best Local Similarity 38.0%; Pred. No. 2.8e-64;
Matches 257; Conservative 145; Mismatches 244; Indels 31; Gaps 12;

Qy 31 AVDLGNRKLEISGKLAFADGASAVVSGDTAVMTAVSKTKRPSQ-QFMPLVVDYROKA 89
Dh 8 SVTMNGOVLLESTGKIRKQADGAVTYKMGNSVLLCTRANVANKAKKEGFLPTINTREMA 67
Qy 90 AAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGNPDTLAI 149
Dh 68 YAAKIPDGFPFKGKASDRVLSRLIDRPVPLFPBGFVNEVAVITVSVANQVNPDIYAM 127
Qy 150 VLAINGVALSLSDIPMNGPVGAVRIGITDGEVYVNTKREKSSSTLNLVAGAPKQIYMLE 209
Dh 128 ILAIIIGASVALSLSDIPMNGPVGAVRIGITDGEVYVNTKREKSSSTLNLVAGAPKQIYMLE 186
Qy 210 VMEASAKENITQDFCAIKYGVKTTQOIIQOIVVETGVTKTKTPQKLTPSPSEIYK 269
Dh 187 MWVESEAHLLSECOMLEAVKFGFSPQVIKIIKEIAEAKKLEWQALYPS--LKGE 244

```

Db      214 SEABLLSDQMGMGAVFPHGEGQQVVIQININELVKBG-----KPRWMDQBPPEVNEALNARV 269
Qy      271 HCLAMERL---YAVFTDYHDKVSDBEAVNKRILDTBEOLKEKFEADPYEIIIESFNVA 327
Db      270 AALAEARLSDAARI-----TDKOERYAOVDVISEITATILABEDTLEDEBELGEIHAIE 324
Qy      328 KEVPSIYILNEBKRCGGRDLTSLRNVSCVDMEKFTLHGSALEFORGOTVLTCTYFDSLES 387
Db      325 KNVARSRLVAGSPRIDGRKEMDIRGLDVRGTGVLPRHGSALLFTRGEOALVTATLGT---- 381
Qy      388 GIKSDQVITAINGIIKDKNFMILYEPFPYATNEIGKVTGLNRREIIGHALAEKALYVPIT- 446
Db      382 -ARDQOVLDELMEGERTDFTLFHYNFPYSVGEGTGMGSPKRRREIGHRLAKRGVLAWMPD 440
Qy      447 -RDPFETIRYSEVLESNGSSMASACGSIALMDGVPRISSAAGVAILGVITKTDPEKG 505
Db      441 MDKFPYTVAVVSEITESNGSSMASVCGASLALMDGVPRIKAAVAGIAMGLVKEG----- 496
Qy      506 EIEDYRLTLDIGIEDYNDMDPKIAGTNGKITALQADIKLPGIPIKIVMEAIQOASVAK 565
Db      497 ---NYVVLSDIDGEDHLDMDPKVAGSRDGISALQMDIKIEGITKEIMQVALLNOAKGAR 553
Qy      566 KEILOINNKITSKPRASRKENGPNVETVOVPLSKAKFVPGGVNKLKLOAETVATISQV 625
Db      554 LHILGVMEQOALNAPRGDISFPARHTTIKINPKIKVDIVSGGSVITALEBTGTITIEIE 613
Qy      626 DEETPSEVAPTPSVNHEARDFTIEICKDDEOOLFEFAGVYATITTEIRDTGVMVLYENM 685
Db      614 DDGTVKIALATDEKAKHAIIRITEI-----TAIEVIGRVYTGKATRIIVDFGAFVAGGK 668
Qy      686 TAVLIANTQOLNBERL 700
Db      669 EG-LVHISQIADKRV 682

```

RESULT 9

E91134
polynucleotide phosphorylase [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E91134
R:Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G.
Gasegawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
A:Reference number: A59629; M01D:21156231; PMID:11258796
A:Accession: E91134
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-734 <HAY>
A:Cross-references: UNIPROT:O8X9W3; GB:BA000007; PIDD:BA837468.1; PID:g13363518; GSPDB:G
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetic8:
A:Gene: EC94045
A:Superfamily: polynucleotide nucleotidyltransferase

[illegible]

Db 214 SEBELLEDOMLGAUVFGHEOOVAYQININELYKAG----KPRDMDPEEVENALMARV 2658
QY 271 HKLAMERL---YAVFTDYEDHKVSRDEAVNKIRLDTBOLKEKPEADPYEIIESFNVA 3272
Db 270 AALAEARLSAPYRI-----TDKQERYAQVDVIXSETIATLLAEDETDENELGELHAIE 3242
QY 328 KEVPRSTVYNEHYRCGRDOLTSIRNVSCEVMDPKTLHLSALFQGGQYOVLCYTFBDES 3873
Db 325 KNVRSRVVLAGEPRIDGREDKIMRGJLDVTVGLPRTHSALFTYGEYDVALVTALTLGT--- 3813
QY 388 GIKSDQVITAINIGIKDKNFMILHYEPFPYATNEIGKVTGILNRRELGHGALAEKALPYIP- 4466
Db 382 -ADAQVLDLDMGERDITFLFHNPFPPYVGEITGVNVS PKRREIIGHGLAARGVLAMPD 4400
QY 447 -RDPFTIRVTSVLESNSSSSMASACGSLALMDGVPPISSAVALGILVYTTDEPKG 5055
Db 441 MDEFPYTVRVAESIETSSSSSSMASVCSGSLALMDAGVPIKAAVAGIAMGLVEKGD---- 4966
QY 506 EIEDYRLNDIIGIEDYNDMDPKTAGNKGTALQADIKLPGRPIKIVMAILOASVAK 5655
Db 497 --NYVALSIDDEHLDMDPKVAGSDGISALQMDIKIEGTKEIMQVALNQAGAR 5533
QY 566 KEIIQJINKTKISPRASRKENGVPVETVOPISPKAKFVGPQYVNLKQLQAEVTGISQV 6252
Db 554 LHLIGVMEQAINAPREDISEFAPRIHTIKYINDKIKVDYIGKGSVIRALTBETGTIIEB 6133
QY 626 DEETFSVAFPTPSVMHEARDPITEIKCKDQEOOLEFGAVYATATITEIRDTGVNWKLYPNM 6655
Db 614 DDGTVKIAATDEGEKAKHAIIRIEI-----TAIEIVGRYTGKVTYRIVDFGAFVAGGK 6688
QY 686 TAVLNLHTQDNERL 700
Db 669 EG-LVHISQJADKRV 682

```

RESULT 10
poly(nucleotide phosphorylase - Deinococcus radiodurans (strain R1)
G75320
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: G75320
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mc-
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:2003696; PMID:10567266
A:Accession: G75320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <WHI>
A:Cross-references: UNIPROT:Q9RSR1; GB:AE002042; GB:AE000513; NID:g6459848; PIDN:AAF160
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2063
A:Map position: 1
C:Superfamily: polyribonucleotide nucleotidyltransferase

Query March 32.5% Score 1156; DB 2; Length 810;
Best Local Similarity 38.1%; Pred. No.4.5e-64;
Matches 261; Conservative 138; Mismatches 236; Indels 50; Gaps 14;

QY 34 LGNRKLTSSGKLFADGSAVVOGDTAVWYTAWSKTPSPGFMPPLVVDYRQKAAAG 93
Dd 41 LGGELSLSTETGKLAKLVGSGVTVRGDTLLVTAQASPTQSKDLPFLTVFEERHYAVG 100
QY 94 RIPTNYLAREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLGNLLAVDGVNEBDVLA 153
Dd 101 KIPSGFQRRGPRGPKAILTSARITDRQIRPLFPKGYRHEQVITVLVSADQGNAPDVLGP 160
QY 154 NGASVALSLDIPMNGPVGAVRIGIIGDEYVYVNPFRKMSSTLTNLVYAGAPKQIVMLE 213

```

```

Db      161 IGAAAALASISDI PMWAPRTACVRVGGQIDGCVVNPPTTEQLFRSRMDVLVAGT-REAVNMVVE 219
Qy      214 ASAENILQODFCHAIKVGVKYYTQITQITGIGIQLVKENG-----VTXKTPQKLFPTSPSEIV 267
Db      220 CCAQCTVSEBDLGVALIEFAHAEMQGVITALIEQMRAEVGHEKENFNLAEBCGPANDVYP----- 274
Qy      268 KYTHLAMERLY-AVFTDYEHDKVSRDEAVNKIRL-----DTEQLMEKPEAPDPEYLIE 321
Db      275 ELTERAKKAAGLDALTTGGKDRSARKLKALRNGULIEGYVDPTPABGSAELTQA---LKD 330
Qy      322 SFENVAKVEFERSI VILNEYKRCGRDLTSLRNVSCEYDMFKTLHGSALEFORGOTVLCTVT 381
Db      331 AAGKVEKRELRRLILLENLRADGRDSKTVRPIVIEARPLPTAGASAVFTGETGVLTGVT 390
Qy      382 FDSLESGIKSDOV---ITAINGIKDKFMLEYBPFPYATNEIGKVTGLNLRREIGHGALA 437
Db      391 L-----GTRDEILIDDLTAAESGDK--FLAHNFPSPYSGEYVWRKMGCOSRREIGHGULA 442
Qy      438 EKALYVPIE--KDFPFTITVTSLEVSNSSSSMAWACGSLAMDGCVPLISSAAGVALG 495
Db      443 KKAIRAVLPSFEFPVIRVAGDVLSENGSSSGTCVACGLTSLMDAGVPLAIPAAGVAMG 502
Qy      496 IYTKTDPKEGJIEDVYLLFTDILGIEDYNGMDPKIAGTNGKIGTALQADIKLPJPIKIM 555
Db      503 LVMEGD-----NRYVLTDILGIEDALGDMDKVCSTAGVTRLQMDIKVGGITTPQIMR 555
Qy      556 EAIQQAASVAKKEILQIMNKTIISKPRASRKENGVEVETOVPLSKRAYVGBGYNLKKLQ 615
Db      556 EMLAQAKBERHLHIGMAEVLAPRAELSPTAHILSLINKPELIGKVIIGGGQVRELE 615
Qy      616 AETGVITISCVDETFESVFAPTPSVMEHARDFTIEICDKDQEQULEFGAVYTAITTEIRDT 675
Db      616 A-MGAQVLTIEBQGVRIFFASGESAFAVKARIEAVTK-----EAKVGBEPBGVIVKIAFP 669
Qy      676 GYMVVLXYPMTAVLLHNTOLDNERL 700
Db      670 GAFVNLFPQDQ-MLHISQLSEQRV 693

```

```

RESULT 11
S74509
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Synechocystis sp. (6
N/Alternate names: protein slll043
C/Species: Synechocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #ext_change 09-Jul-2004
C/Accession: S74509
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yaoda
DNA Res. 3, 109-136, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
B.
A/Reference number: S74322; NCID:97061201; PMID:8905231
A/Accession: S74509
A>Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A.Residues: 1-718 <KAN>
A.Cross-references: XANPROT:P72659; EMBL:D90899; GB:A001339; NID:g1651650; PIDN:BA1666;
A.Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A:Gene: pnp
C:Superfamily: polyribonucleotide nucleotidyltransferase
C:Keywords: nucleotidyltransferase

Query Match      32.4%; Score 1153; DB 2; Length 718;
Best Local Similarity 37.5%; Pred. No.5.8e-64;
Matches 257; Conservative 131; Mismatches 261; Indels 36; Gaps 10;

Oy      31 AVLDGNRKLEISGKLARFADGSAAVQSGDTAVMTVAATKTKSPSQFMPLVVYRCKA 90
       ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      7 SISFGDRDIRLRMGTLAPAGSGSVLLQSGETVALVTATRAKGRDGFPLPVTVDYEGRLX 66

Oy      91 AAGRIPNTALREVGTSDEILSTRILIINRSIRFLPFAGIYYTDVQLCNLAADVNEEDV 150

```

```

Db      67 AAGRIPGGFLRREGGPPREKATLISRLIDRLPLRPLPPLMLRDELQVATTLMSDEEVPDV 126
Qy      151 LAINGASVALSLSDIPMNGPVGAVRIGIIDEEYVNPFRKEMSSSTLNLVVAAGPKSQIV 210
Db      127 LAVGASVAVILAIQIPFKGPMAAVAVGVGDPIINPYREVNNDLILVVAAGT-AGIV 185
Qy      211 MLEASAEVILIQODEFCHAIKGVKTYQOIIQGLQVKEGTGVTKTPQKLFPPSP---EIV 267
Db      186 MVEAGANLPPRODIIIEALDFGEAVQDILINARELMTLGLTILATSE----PPVNTAVE 241
Qy      268 KYTHKLMERLVAFTDYEHDKVSRDEAVNKIRLDTESQLKEKPEADPYE----- 318
Db      242 EFINARSKKIITVLGDFDLGMDGDAALDEIKATEVETALAELETPDVAKSVSEDPKL 301
Qy      319 IIESFNVAKEVFSIVNEVKRCGROLTSLRNVCSEVDNF-KTLHGSALFORQTOVL 377
Db      302 VGNLYKALTKKLMRKQIVDDGVRRVGRKLEQVRPISCEVGLPRRVHSGGLFNRGLTQVL 361
Qy      378 CTVPFDSLSESGIKSDQVITAINGIKDKNFMHYEPFPATNIEIGKVTGLNRELGHALA 437
Db      362 SLAT---LGSFGDADLDLHPEDEKRYLHHYVFPYISVGEARPMRSPGRREIGHALA 418
Qy      438 EKALYVPIP--RDEPFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAG 495
Db      419 ERATIPVLPQEDPEPVYVAVSVLENSGSGTSLALMDAGVPIKKPVGALMG 478
Qy      496 LVTTDEPKGEIEDYRLTLGLIEDVNDMDFKAGTNKIGITALQADIKLPGIPIKIVM 555
Db      479 LIKGD-----BIRILITDQIGIEDFLGDMDFKAGTDSGITLALQDMKIDGLSMERVS 531
Qy      556 EAIQOASVAKKEIIOIMNKTISKPRASKENGPPVETQVPLSKAKVFGGVNLKTLQ 615
Db      532 KALQALPARKHILDKMLATIRREPPRLSPAPRILITKIEBENIGAVIGEGKTIKIT 591
Qy      616 AETGVTSIQVDETFSEVAPTPSVMHARDFTEICDQOEQLEFGAVYATITIEIRDT 675
Db      592 EGTSCCKIDIDAGVTITASSGSEBRAROMIYNTMR-----KLNESGVYGRVTRILPI 646
Qy      676 GVMVKLYPNMTVALTLNLTQDNERL 700
Db      647 GAFVEVLFGKEG-MIHISQLTEGRV 670

RESULT 12
H65106 polynucleotide nucleoside diphosphate transferase (EC 2.7.7.8) alpha chain - Escherichia coli (E
N:Alternate names: polynucleotide phosphorylase
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence revision 23-Jan-1998 #text_change 09-Jul-2004
C:Accession: H65106; A26118; E23984
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.D.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65106
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-734 <BLAT>
A:Cross-references: UNIPROT:P05055; GB:A000397; GB:U00096; NID:G2367199; PIDN:AAC76198.
R:Experimental source: strain K-12, substrain MG1655
R:Regnier, P.; Grunberg-Manago, M.; Potter, C.
J. Biol. Chem. 262, 63-68, 1987
A:Title: Nucleotide sequence of the pnp gene of Escherichia coli encoding polynucleotide
in S1.
A:Reference number: A26118; MUID:87083499; PMID:2432069
A:Accession: A26118
A:Molecule type: DNA
A:Residues: 'M', 25-379, 'R', 381-472, 'S', 474-734 <REG>
A:Cross-references: EMBL:J02638; NID:G147744; PIDN:AAA83905.1; PID:G551833
R:Evans, S.; Dennis, P.P.
Gene 40, 15-22, 1985

```

```

A:Title: Promoter activity and transcript mapping in the regulatory region for genes enc
A:Reference number: A23984; MUID:86137413; PMID:3005122
A:Accession: B23984
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 'M', 25-85 <EVA>
A:Cross-references: GB:M14425; NID:G147747; PIDN:AAA24596.1; PID:G147749
A:Note: the authors translated the codon GCC for residue 74 as Arg
C:Genetics:
A:Gene: pnp
A:Map position: 69 min
A:Start codon: TTG
C:Superfamily: polynucleotide nucleoside diphosphate transferase
C:Keywords: nucleoside diphosphate transferase
F:25-734/Product: polynucleotide nucleoside diphosphate transferase alpha chain #status predicted

Query Match      32.4%; Score 1152; DB 2; Length 734;
Beat local similarity 39.9%; Pred. No. 6.9e-64;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;

Qy      35 GNRKLEISGGKLARPADSAAVVGSGDPTAVMTAVSKTPSPSO-FMPLVVDYRQKAAAG 93
Db      35 GQHTVLTETGMARQATTAAVVMSMDTAVFVTVVGQKAKGQDFPLTVNYQERTYAG 94
Qy      94 RIPTNYLRREVGTSDKEITLRIIDRSIRPLFPAGYFDTQVLGNLAVDGVNEPDLAI 153
Db      95 RIQSPFRREGRPEBGETLIRLIDRPIRPLFPFGFVNEVVIATVSVNVQNVADIVAM 154
Qy      154 NGASVALSLSDIPMNGPVGAVRIGIIDEEYVNPFRKEMSSSTLNLVVAAGPKSQIVMLE 213
Db      155 IGASALSLSGIPFNGPIGARVGVINDQVYLANPQOELKESKIDLVAAGT-EAVALMVE 213
Qy      214 ASAENILQODFCHAIKGVKTYQOIIQGLQVKEGTGVTKTPQKLFPPSP---EIVYKT 270
Db      214 SEAOQLSEDMGLGVAVFGEHQOQVQVQINLVLEAG---KPRMDQPEEVNENALNARV 269
Qy      271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTESQLKEKPEADPYEIIISFNVA 327
Db      270 AALAEARLSDAYRL-----TKQERYAADVYKSETIATLAEDETLENELGELIAIE 324
Qy      328 KEVERSIYLVNKKYKCDGDLTSLRNVCSEVDNMFTELHGSALFORQTOVLCTVPFDSL 387
Db      325 KNVVRSRLVAGEPRIDGSEKDMIRGLDVRGVLPRTGSALEFTRGETQALVTATLGT--- 381
Qy      388 GIKSDQVITAINGIKDKNFMHYEPFPATNIEIGKVTGLNRELGHALAKALYVPIP- 446
Db      382 -ARDAQVLDLIMGERTDTFLPHYVFPYISVGETGVSGPKRREIGHGLARGVLAAMPD 440
Qy      447 -RDEPFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAGVLTITDPEKG 505
Db      441 MDKPPYTRVRSVSTTESNGSSSMASVCGASLALMDAGVPIAAVAGIANGLVKGD---- 496
Qy      506 EIEDYRLTLGLIEDVNDMDFKIAGTNKIGITALQADIKLPGIPIKIVMAIQOASVAK 565
Db      497 ---NYVVLSDILGDEHLDGMDFKVAGSRDGSIALQMDIKIEGTKEIMQVALNDKAGAR 553
Qy      566 KEIIOIMNKTISKPRASKENGPPVETQVPLSKAKVFGGVNLKTLQAEVTSIQV 625
Db      554 LILIGVMEQAINAPRGDISPPAPRIHTIKINDIKOVIGGGVIRALTBETGTIIE 613
Qy      626 DEETFSVAPTPSVMHARDFTEICDQOEQLEFGAVYATITIEIRDTGVMVKLYPNM 685
Db      614 DGGTVKATATGGEAKAIAIRIEI-----TAIEVGRVITGVKTRIVDFGAFVAGGK 668
Qy      686 TAVLHNTQDNERL 700
Db      669 EG-LVHISQIADKRV 682

RESULT 13
AC3497 polynucleotide nucleoside diphosphate transferase (EC 2.7.7.8) [imported] - Brucella melitensis
C:Species: Brucella melitensis

```

C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C/Accession: AC3497
R/DelVecchio, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
P.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A/Reference number: AD3252; PMID:11756688
A/Accession: AC3497
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-714 <KUR>
A/Cross-references: UNIPROT:Q8YEB7; GB:AE008917; PIDN:AAL53142.1; PID:G17984011; GSPDB:C
A/Experimental source: strain 16M
C/Genetics:
A/Gene: BMEI1961
A/Map position: 1
C/Superfamily: polyribonucleotide nucleotidyltransferase
C/Keywords: nucleotidyltransferase

Query Match 32.3%; Score 1150; DB 2; Length 714;
Best Local Similarity 39.0%; Pred. No. 8.8e-64;
Matches 268; Conservative 124; Mismatches 247; Indels 48; Gaps 14;
Qy 30 VANDLGRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYRQK 88
Db 7 VEIEMGRPLTLETGKIARQADGAVLATYGETAVLATVSAKEPFGDPFPLTVNYQEK 66
Qy AAAAGRIPTNVLREVGSDKEILTSRIIDRSIRPLFPAGFYVDYQVLCNLAVDGVNEP 148
Db 89 AAAAGRIPTNVLREVGSDKEILTSRIIDRSIRPLFPAGFYVDYQVLCNLAVDGVNEP 148
Qy 67 TYAAGKIPGGYFKKGRSEWETLVSRIDRIPRLPFDGKADTQVITVYVQLQDLENNP 126
Db 149 DVLAINGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQ 208
Qy 127 DILSMVASALTLTISGVPFPGISGARVYIDGEVLPNDENPESKLDLVVAGTSA 185
Db 209 IWMLEASAEENLQODFCHAIKVYKTYQIIQGIQOLVKEGVTYKRPQKLFPSPELVK 268
Qy 186 VLMWSESEOLPEBVMGAVMFGHKSFPQVIDAI--IKLAEVAKEPRDF--QPEDLS 239
Db 269 YTHKLAMERLAVATDYEH-----DKVSRDEAVNKIRLDTEBQKEKPE-----AD 315
Qy 240 ---ELEAKVLAVENTDLREAVKITEKQRIYAAVDAKAKAEHF--PBEVETENSAA 293
Db 316 PYEIESFNVAKEVFSIVLNEYKCDGRDLSLRNVSCEVDMFKTLHGSALFQSGQTQ 375
Qy 294 QFATL--FKHLQAKIVMNIIDTGNRIDGRDLSLRNVSCEVDMFKTLHGSALFQSGQTQ 351
Db 376 VLTVTYDLSGSGIKSDQVITAINGIKDKNFMHYEPPPYATNEIGKVTGLNRRLGHGA 435
Qy 352 AIVVATLTGTGE---DEQMDALGTYESFMLHYNFPYVSGETGRMGSPGRREIGHGK 407
Db 436 LAEKALVPVIR--RDPPTIRVTSEVLSENGSSMASCGSLALMDSGVPISSAVGVA 493
Qy 408 LAMRAIHMLPAQOFPTTIRAVSELITESNGSSMAVTCGSLALMDGVIIVAPVAGIA 467
Qy 494 IGLVTKTDPEKGEIEDYRLTDIIGIEDYNDMDFKIAGTKGITALQADIKLPGIPIKI 553
Db 468 MGLIK-----EERFAVLSDIIGDEBHLGDMDFKAAGTEFGTSLQMDIKIDGITEEI 520
Qy 554 VMEAIQOASVAKKEILQIMNTKTIKSPRAKRENGPVVETVPLSKRAKFGPGYNLKK 613
Db 521 MKVALQAKGGRVHILGEMAKAIISSRAELGEFAPRIEVMNIPTDKIRDIVGSGGKAVRE 580
Qy 614 LQATGTGISQVDEETPSVFAPTPSVMEHADPTTEICKDQEOOLEGFAVYTAITIEIR 673
Db 581 IVEETGAKINIEDGTAKIASSNGKEIEAAKKWIHSIAEP-----EVEGTYEGIVVXTA 635
Qy 674 DTGVAVKLYPMNTAVLLHNTQLDNERL 700
Db 636 DFGAFVNFPGPRDG-LVHISQLAADRV 661

RESULT 14

AC0424
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - *Yersinia pestis* (str:
C/Species: *Yersinia pestis*
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C/Accession: AC0424
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; F
11, M.; Rutherford, K.; Skellern, J.; Skellern, J.; Skellern, J.; Skellern, J.; Skellern, J.;
Nature 413, 523-527, 2001
A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: AC0424
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-705 <KUR>
A/Cross-references: UNIPROT:Q8ZBC6; GB:AL590842; PIDN:CAC92719.1; PID:G15981414; GSPDB:G
A/Genes: pnp
C/Superfamily: polyribonucleotide nucleotidyltransferase
C/Keywords: nucleotidyltransferase

Query Match 32.2%; Score 1146; DB 2; Length 705;
Best Local Similarity 40.0%; Pred. No. 1.5e-63;
Matches 263; Conservative 113; Mismatches 259; Indels 32; Gaps 10;
Qy 35 GNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYRQKAAAAG 93
Db 12 GQHTVTIETGMMARQATAAVVMSMDTAVFVTVVQCKAKPGSFFPLTVNYQERTYAG 71
Qy 94 RIPNRYLRREVGSDKEILTSRIIDRSIRPLFPAGFYVDYQVLCNLAVDGVNEPDLAI 153
Db 72 RIPNRYLRREVGSDKEILTSRIIDRSIRPLFPAGFYVDYQVLCNLAVDGVNEPDLAI 131
Qy 154 NGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQIYMLE 213
Db 132 IGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQIYMLE 190
Qy 214 ASAEENLQODFCHAIKVYKTYQIIQGIQOLVKEGVTYKRPQKLFPSPELVKYNH-- 271
Db 191 SEADILSEBQMLGAVVGHGQGVLEINALVAEAKPMDWQ-----AEVNEALHAR 245
Qy 272 --KLAMERLAVATDYEHDKVSRDEAVNKIRLDTEBQKEKPEADPEYIEESFNVAKE 329
Db 246 VAEIAERLIDAKRITE--KQERTYQVDALKADVTETLAAQDDTLDAEIIQDILASVEKN 303
Qy 330 VFSIIVLNEYKCDGRDLSLRNVSCEVDMFKTLHGSALFQSGQTQVLTVPDLSGSI 389
Db 304 VMSKRVLRGPRIDGRKMDIRGIDVARTGILPRTHGSALFTREGTQALVTATLGTAADQ 363
Qy 390 KSDQVITAINGIKDKNFMHYEPPPYATNEIGKVTGLNRRLGHGALAEKALYFVIP--R 447
Db 364 NIDELM-----GERDSFLHYNFPYCVGETGMGSPKREIGHGRILAKKGVLAVMPSAS 419
Qy 448 DFPPTIRVTSEVLSENGSSMASCGSLALMDSGVPISSAVGVALGVTKTDPEKGEI 507
Db 420 EFPPTIRVTSEVLSENGSSMASCGSLALMDSGVPISSAVGVALGVTKTDPEKGEI 473
Qy 508 EDYRLTDIIGIEDYNDMDFKIAGTKGITALQADIKLPGIPIKIYMEAIQOASVAKKE 567
Db 474 -NEVVLSDIIGDEBHLGDMDFKAAGTEFGTSLQMDIKIDGITEIEDD 532
Qy 568 ILQIMNTKTIKSPRAKRENGPVVETVPLSKRAKFGPGYNLKKLQATGTVTISQVDE 627
Db 533 ILGVMEQALISTPDDISEFAPRIYTKINPEKIKDVIKGGGSVALRLDTDEGTTEIEDD 592
Qy 628 ETSVEVPAPTPSVMEHADPTTEICKDQEOOLEGFAVYTAITIEIRPTGVAVKLYPMNTA 687
Db 593 GTIKIAATTDGAKKHAIRIEEI-----TAEIEGRIVAGKYRIYDFGAFVAGGKEG 647
Qy 688 VILHNTQLDNERL 700
Db 648 -LVHISQLAADRV 659

RESULT 15

883950
polynucleotide phosphorylase pnpA [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: G83950
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83950
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-704 <STO>
A:Cross-references: UNIPROT:Q9KAB3; GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA061
A:Experimental source: strain C-125
C:Genetics:
A:Gene: pnpA
C:Superfamily: polynucleotide nucleotidyltransferase

Query Match 32.2%; Score 1145; DB 2; Length 704;
Best Local Similarity 39.3%; Pred. No. 1.8e-63;
Matches 265; Conservative 132; Mismatches 251; Indels 26; Gaps 11;

Qy	31	AVDLGNRLKLSISGLARFADGSAVQSGDTAVVWTAVSRTSPSQFPLVVDYRQAA	90
Db	9	SIDWAGKLTVTCTQLAKONGAVLVRYGTAVLSTATASKEPKDLPFPPLTVNYEERLY	68
Qy	91	AAGRIPTNYLAREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEBDV	150
Db	69	AAGRIPTNYLAREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEBDV	128
Qy	151	LAINGASVALSLDIPNNGPVGAVRIGIDSEYVNPYRKEMSSSTLNLVYAGAPKQIY	210
Db	129	AAWVGSSIALISDIPPEGPAGVAGIDQFVINFQDLKESDIHLVVAAGT-KDAIN	187
Qy	211	MLEASANILOQDFCHAIKVGVKTYTOIIGIQQLVKSGVTKRTPQKLPFSPEIYKYT	270
Db	188	MVEGAEVEVEDVNLKALMEGHNEIKRIEPOEKIAAEVGSK-TDVLTKQVDMLEQEV	246
Qy	271	HKLAMERL-YAVFTDYEHDKVSRDEAVNKLRLDTEQLKEKFPADPYEIIIESFNVAKE	329
Db	247	RVKAEEDLKQAVQVPEKARQDAIEAVMDKVLFTYED-NEDVPLSEVNEILHK--LYKE	302
Qy	330	VPSISVLAKEYKCCGRDLTSIRNVSCVDMMKTLHGSALPFGQTOVLCTVTPDSLSGI	389
Db	303	EVRLITVEKIRPDGREIDIRPLSSQVILPRTGSGLPFRGQTOALSTCTGAL----	358
Qy	390	KSDQVITAINGIKDKNFMHYEPPTVATNEIGKVTGLNRRRLGHGALAEKALYVIP--R	447
Db	359	GDVQIILGIGIEESKRFMAHYNFPQFSVGETPIRGQREIGHGALGERLLEVPIDSEQ	418
Qy	448	DPEPTIRVTEVLESNGSSSMASACGSLALMDSGVPISSAVAGVALGLVTKTDPKEGEI	507
Db	419	DPEPTIRVTEVLESNGSSSMASACGSLALMDSGVPISSAVAGVALGLVTKTDPKEGEI	471
Qy	508	EDRRLTLIDIGIEYNDMPKIKTGKITALADIKLPGIPIKIYMEAIQASVAKGE	567
Db	472	EHVSVLTDIOGMEDALDMDPKVAGTRKGYTALQMDIKISGIDRALILEQALEQARKGMI	531
Qy	568	ILQIMNTISKPPRSRKENGPNVETVOVPLSKRAKFPVGGGVNKKLQAEFGVTISQVDE	627
Db	532	ILDNMLEAISERSSELSFPYAKILMTINPKIRDVIGPSGKIMNKIIBDTGVKIDIBQD	591
Qy	628	ETFSVPAPTPSVHAEADPITIECKDQEQQLERGAVALYATITIEIRDTGVVVKLYPNMTA	687
Db	592	GTIYISSADTVMNKKABEIIEDIVR-----EVEVGQVYLGTVKRIEKGAFVELFKGKDG	646
Qy	688	VLAHNTQLDNERLN 701	
Db	647	-LVHISQLAEBRVN 659	

Search completed: January 28, 2005, 19:35:42
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using SW model

Run on: January 28, 2005, 19:28:42 ; Search time 40 seconds
(without alignments)

1168,855 Million cell updates/sec

Title: US-09-907-907a-42

Perfect score: 3557
Sequence: 1 DGFPLPRDRALTLQVRA.....TAVLLNTQDNERLILP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 6631800 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1165.5	32.8	706	US-09-134-001C-4908	Sequence 4908, App
2	1150.5	32.3	719	US-09-489-039A-13639	Sequence 13639, A
3	1140	32.0	720	US-09-543-681A-5738	Sequence 5738, App
4	1129.5	31.8	708	US-09-134-000C-5994	Sequence 5994, App
5	1125	31.7	716	US-09-809-665A-171	Sequence 171, App
6	1107	31.1	750	US-09-107-532A-5868	Sequence 5868, App
7	1103	31.0	699	US-09-340-236-2259	Sequence 2259, App
8	1101.5	31.0	684	US-09-809-665A-109	Sequence 109, App
9	1065	29.9	705	US-09-328-352-8112	Sequence 8112, App
10	1053	29.6	737	US-09-583-110-2924	Sequence 2924, App
11	976.5	27.5	697	US-09-252-991A-27283	Sequence 27283, A
12	884	24.9	566	US-09-198-452A-1073	Sequence 1073, App
13	860.5	24.2	332	US-09-270-767-44788	Sequence 44788, A
14	710	20.0	488	US-09-809-665A-44	Sequence 44, App1
15	636	17.9	358	US-09-809-665A-155	Sequence 155, App
16	389.5	11.0	247	US-09-710-279-2006	Sequence 2006, App
17	340.5	9.6	91	US-09-270-767-60251	Sequence 60251, A
18	226	6.4	127	US-09-198-452A-1074	Sequence 1074, App
19	175	4.9	52	US-09-513-999C-6411	Sequence 6411, App
20	173	4.9	56	US-09-513-999C-7240	Sequence 7240, App
21	144	4.0	455	US-09-134-000C-6074	Sequence 6074, App
22	140.5	3.9	1233	US-09-134-000C-4971	Sequence 4971, App
23	140.5	3.9	1416	US-09-071-035-404	Sequence 404, App
24	140.5	3.9	1448	US-09-071-035-402	Sequence 402, App
25	140	3.9	252	US-09-252-991A-23776	Sequence 23776, A
26	139	3.9	921	US-09-800-729-199	Sequence 199, App
27	138	3.9	119	US-09-270-767-60250	Sequence 60250, A

28	137	3.9	191	4	US-09-252-991A-27186	Sequence 27186, A
29	136.5	3.8	1575	4	US-09-917-254-83	Sequence 83, App1
30	136	3.8	257	4	US-09-543-681A-7481	Sequence 7481, App
31	132	3.7	828	4	US-09-540-236-3107	Sequence 3107, App
32	129	3.6	259	4	US-09-489-039A-10604	Sequence 10604, A
33	128.5	3.6	244	4	US-09-328-352-5628	Sequence 5628, App
34	127.5	3.6	871	3	US-09-134-001C-3979	Sequence 3979, App
35	124.5	3.5	471	4	US-09-107-532A-5705	Sequence 5705, App
36	124.5	3.5	1306	3	US-08-999-774A-13	Sequence 13, App1
37	123.5	3.5	868	4	US-09-800-729-106	Sequence 106, App
38	118.5	3.3	1781	4	US-09-961-403-13	Sequence 13, App1
39	116.5	3.3	246	4	US-09-540-236-3210	Sequence 3210, App
40	116	3.3	1780	1	US-08-769-309A-5	Sequence 5, App1
41	116	3.3	1780	3	US-08-994-570-5	Sequence 5, App1
42	115	3.2	875	4	US-09-107-532A-4537	Sequence 4537, App
43	114	3.2	648	4	US-09-198-452A-594	Sequence 594, App
44	114	3.2	2270	4	US-09-581-909-3	Sequence 3, App1
45	113	3.2	235	4	US-08-747-562-28	Sequence 28, App1

ALIGNMENTS

RESULT 1
US-09-134-001C-4908
Sequence 4908, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4908
LENGTH: 706
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4908

Query Match	33.8%;	Score 1165.5;	DB 3;	Length 706;
Best Local Similarity	38.7%;	Pred. No. 2.4e-100;		
Matches	270;	Conservative 123;	Mismatches 250;	Indels 55; Gaps 13
QY	24	SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVOGDTAVVTVAVSKTKPSQFMPLVV	83	
DB	7	SQEKVKTETWAGRSGLTERTGQLAKQNGAVLVRYGTVVLSRTAKERRDGFPLTV	66	
QY	84	DYRKAAAGRIIPNYLRREVSTGDEILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVD	143	
DB	67	NYEKVYAAAGKIPGFKRGREGRPDEBATLTARLIDRIRPLFPKGYHVDQIMNIVLSAD	126	
QY	144	GVNEPDTLNGASVALSLSDIPNGRGVAGVRIIGLGEVNVNPTREKMSSTLANLVAG	203	
DB	127	PDGSEPMAMAGSSMALIVSDIPQGIAGVNVGYDGKVIIVNSVADKEISRDLVLVAG	186	
QY	204	APKSOIYMLEASENILQODFCHAIKVG-----VKYTOIIOGLOLVKETGVTRTP	256	
DB	187	HKDVAVNVVEGASEITESEMLAIFGHEEIKRLVAFQGEIIDHIQPIQE-----	237	
QY	257	QKLTPT---SPETVYKTKLAMER--LVAVFTDYEHDKVSRDAVNRKRLDTBOLKEKF	311	
DB	238	---FVPERDEDLVEKVKSLTBEDGLKDVLT---FDKQQRDEMLDLAK---EEVVGHL	288	
QY	312	PEADP-----YEISFNVAKEVPRSYLVNRYKRCDDGDLSTLRVSGEVDMLKHLG	365	
DB	289	DEEDPENETLVKEVYALINDLKEEVRLTADKIRPDGKRVKDEIRPLESEVGLPRAHG	348	

Qy	366	FALPQFGTOVLCVTPDPSLESGISDQVITAINIGIKDQNFMLHYEFPYATNTEIGKVNG	425
Db	349	SGLFTRGTQALSVTLTGAL-----GDYQLIDELGEVEVKRPMHYNFPMFVSGETGPVYA	404
Qy	426	LNRRRLGALAEKALVPVIP--RDPEPTIRVTEVLESNGSSSMASACGSLIAMDGV	483
Db	405	PGREIRIGGALGERALRYIIPDTQFPFYIRIVSEVLEBSNGSSSQASISGSLTALMDAGV	464
Qy	484	PISAVAVAGLGVKTPDPEKGEIEDYRLTLIIGIEDVNGMDPFIAGTNGGITLALQAD	543
Db	465	PIKAPVAGIANGLVTRD-----SYTLITDIQGMEDLACGMDPRVAGTNGGITLAIQMD	517
Qy	544	IKLPEIPIKIYMEALIQASVAKKEILQIMNKTIISKPRASRKENGVPVETVOYPLSGRAKF	603
Db	518	IKIQDLTEVIEEALAEQARQGRLAIMDNLMTIEQPRELSAYAPCVVTMSINPDIRDV	577
Qy	604	VGPQGYNNKKLOAETGVGISQVDEETFSVFATPPSWHARDPFIETIEICDDOEQOLEFGA	663
Db	578	IGPGSKKINELIETDELGVGLDIEQDGTIFIGADVQAMINAKELIIBDITR-----EAEVQ	632
Qy	664	VYTAITITEIRDTGVWVKYLPNNKTAVLLANTQLDNERLN	701
Db	633	VYHAKVKRIEKYGAVLELFPQKDA--LLHSISQISQEBIN	669

```

RESULT 2
US-09-489-039A-13639
; Sequence 13639, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13639
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13639

```

Query Match	32.3%;	Score 1150.5;	DB 4;	Length 719;
Best Local Similarity	38.2%;	Pred. No. 6.3e-99;		
Matches 270;	Conservative 124;	Mismatches 261;	Indels 51;	Gaps 12;

Qy	6	LPBPRALTOLOVRLALMSAGSRVAVVDLGRKKLEISGOKLRAFDGSAVYOSGDPAWV	65
Db	2	IDKKORILLNPVIRKF-----QYGHVTVLLETGMMAQAAVAWVSMDBPAVF	50
Qy	66	TAVSKTSPSPQ-FMPVLVDVYROKAAAGRIPTNNYLRREVGTSDEKILTSRIIDRSIRPL	124
Db	51	TVGGCKAKKPGODEFPPLTYNYOERTYAACKITCGFPFRGRGRSSEGETTLARLIDREYRPL	110
Qy	125	FPAGFYDYOVLCSNLAAVDGVAEPDVLALNGASVALSLSDIPMNGVGAVRIGIIDGEYV	184
Db	111	FPFGVNVENVQVATVAVSVNPNQVNPDI VNMATIGASALSLSGIPRNGSIGARVGYINDQYV	170
Qy	185	VNPTRKEMSSSTLNLVAVAGAPKSQIVMLFASAEENILQODFCAIKVGVKYYTOIIGSIOO	244
Db	171	LNPTQOELEKSSKLDLNVAGT--EAAVLMVSEAEKLISEDQMLGAVVFGHEQQOIVIQIND	229
Qy	245	LYKEFGVTKRTQKLFPSPEIYK-----YTHKLAMERL---YAVVTYIEHDKVSGDEAV	256
Db	230	LYKEAGKPRMDWQ-----PEAVNEMALNARVAALAESRLSDAYRI-----TDQCEYAAQV	278
Qy	297	NKIRLDTEBQLEKFPPEADPYELISEFNVAKEVFSITYLNEKRCODGDLTSLRNVSC	356
Db	279	DVKESLTATVLAEBEETLDANELGELIHAIEKNVNASRVLAGEPRIDRGEKMDIRGIVR	338

```

QY 357 UDMFTHLGSALFQGOQOVUCLTVFEDSLSEGI KSDOVUATANGIDKMFMLHYEPPPA 416
Db 339 TGVLPRTGSHLFTFGETQALVTALTLGTARADQNIDELM----GERTDSFLFHYNFPYS 394
QY 417 TNEIGKVTGLNREELGHGALAEKALYPYIP-RDPFFITRTVSEVLESNGSSMBAACGG 474
Db 395 VGETMGWSPKRRERIGHRLAKRGVLAMPTEEPPYTVRVSEITESSGSSMAVCCA 454
QY 475 SLALDGSVPFISSAVAGVAILGVLTIDPEKGEIEBYRLTDLIGIEDVGDMDFKIAGTN 534
Db 455 SLALMDADVPKAAVAAGTAMGLVKEGD-----NFVVLSDILGDEBDHLGDMDFKVAGSR 507
QY 535 KGITLQADIKLPGPIKIVMEALIOASVAKKEILQIMNKITSKPAASKENCPVETQ 594
Db 508 DGISLQMDIKIEGITEKIMOVALNOQAGARLHILIGVEQAINAPBGDISFAPRHHITK 567
QY 595 VPLSKRAFVGGGYNLKKLQAEITGVITISQVDEETFSVEAPRPPVVMHEARDPITEICKD 654
Db 568 INPDKIKOVIGKGSVIRALTEBETGTTIEIEDDGTVKIATGDGKAQHARIREEI----- 623
QY 655 QEOQLFEGAVYATITTEIRDTGWMVKLPYNNMTAVULHNTOLNERNL 700
Db 624 -TAEIEGRTINGKVRILVDFGAFYALGGKEG-LVHLSQIDKRV 667

```

```

RESULT 3
US-09-543-681A-5738
; Sequence 5738; Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5738
; LENGTH: 720
; TYPE: prt
; ORGANISM: Proteus mirabilis
US-09-543-681A-5738

```

Query Match	32.0%;	Score 1140;	DB 4;	Length 720;
Best Local Similarity	39.2%;	Pred. No. 6.1e-98;		
Matches 264;	Conservative 123;	Mismatches 254;	Indels 32;	Gaps 11;

```

QY      35 GNRKLEISSGLARAPADOSAVVQSDPTAMWTAASFKTSPSO-FMPLVVDYRCKAAAG 93
Db      23 GQHTLTLEGMARAPATAVWMDMGTAFVTVIAKKVKEGODFFPLTVNIOERBYAAG 82
QY      94 RIFTVYLTREVEGTSDEKILTSRIIDRSIRKPLFPAGFYDTQVLCMLAAVDGNEPDLAI 153
Db      83 RIFGSPFRREGRPGSGEITLIARLIDRPLRPLFPGEGTANIGVATVAVSNPQVNPDIAM 142
QY      154 NGASVALSLSDIPMNGPVGAARIGIIDGEYVNPTRKENSSTLNLVAGAPKSOIVMLE 213
Db      143 IGASALSLSGVPFNGPIGARGVGFIDGQVLANPITDELKISKLDLIVAGT-AGAVIAYE 201
QY      214 ASAENIILQDFCHAIKVGKTKTQOIIQIGQVLKETGVTKRTPQKLFITSPSLVKTKH-- 271
Db      202 SEADLTSEEQMIGAVVFGHEQOQVVIENINALVAEKGKMW----WAEPE-INOGLHNR 256
QY      212 --KLAMEELVAVFTDYEDHKVSRDAVNKIRLDTEBQLEKRPENDPVEIISFNVAKE 329
Db      257 IAGLQTRTIGAVRYITE--KQERYEQIDAIRBEVISTLLAEBSLDEGEIIEIFSGLEKK 314
QY      330 VFRSLVNLMEYRCQDRDLTSLRNWSCVDWMFTLLGSALEFGOGOTVLCVTFPDSLESGI 369
Db      315 IVRAVLVLGEPPIIDREKDMVWALDIRGLRPRTHGSAFTTGEOTDALVTATLGT----A 370

```

```

0Y 390 :SSDDVYITANGIKDKNFMALHYEPPPAATNEIGKVTGLNRELDHGLAELKALPVPR--R 447
Db 371 RDAQITIDIMEBHHTTFLAHINPPYVSGETMGSPKREIEIGHRLKAGVLAHVETIS 430
0Y 448 DEFPRTIRYSEWLEBSNGSSSMASACGSLALMDSGVPIISANAVGVALGLVTKTDPKGEI 5070
Db 431 EFPYVVRVSEITEESNGSSSMASVCCASLALMDAGVPIESVAGIAMGLV-----EG 4833
0Y 508 EDYRLLTDLIGIEDYNGDMDFKIAGTNKIGITALQADIKLPGIPIKIWEALIQASVAKKE 5675
Db 484 ENFVVLSDILGEDHDHLDGMDFKVAGSRNGVSAIQMDIKIEGIRREIMQVLAQAKSARLH 5433
0Y 568 ILOINNKTIISKPRASKEKGVVETVQVPLSKRAKVVGGGVNKLKQLQAEVTYISOVDE 6227
Db 544 ILGVESALISOPRAEISEFAPRIHTIKINADKI KDVIYGKGSVYRALTEBETGTTIBIEDO 6030
0Y 628 ETSFVEAPTPSVMEHARDRITREICDODQOQLEFGAVYATITREINADGTGVNVLXPNMTA 6687
Db 604 GTVKAALATSGEOAKAIAIRIEI-----TAEVGRGIRYNGKVRIVDFGAFVALIGGKEG 6586
0Y 688 VLIHNTOLDNERLT 700
Db 659 LVHISQIADKRV 670

```

```

RESULT 4
US-09-134-000C--5994
: Sequence 5994, Application US/09134000C
: Patent No. 6617156
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
: FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 032796-032
: CURRENT APPLICATION NUMBER: US/09/134,000C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/055,778
: PRIOR FILING DATE: 1997-08-15
: NUMBER OF SEQ ID NOS: 6812
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 5994
: LENGTH: 708
: TYPE: PRT
: ORGANISM: Enterococcus faecalis
US-09-134-000C--5994

```

Query Match	31.8%	Score 1129.5	DB 4	Length 708
Best Local Similarity	36.6%	Pred. No. 5.8e-97		
.Matches	259	Conservative 147	Mismatches 248	Indels 53
Gaps	15			

QY	9	RDRALTLQGV-RALWSSAGSRAVAVDLGNRKLTIESSGKLAREPADSAVVGSDTAVMYTA	67
DB	1	RRRLMTEKQVFKTWM-----GGRPLFEVIGQLAKQANQAVLYRVGDTVYLSAA	48
QY	68	VSKTKPSSQSMPLVVDYRQKAAAGRIPTNYLAREVGTSDKEILTSRIIDSITPLPA	127
DB	49	VASKEAKVDPEFLTVNVEEKMYAVGKIPGCFIKKEGPRSEATLTARLIDRPIRMFSQ	108
QY	128	GYFVDYQLCNLMLNDGVPEDPYLAINGASVLSLSDIPMNGPVCANRIGIIDGVNYP	187
DB	109	SFRNEVQITNINVSVEQDCTPEMAAMFGSSLALASDIPFDSPIAGVDVGRINGEVLNP	168
QY	188	TRKEMSSSTLNLVVGAPKSGQIVMLAEASAEINLODFCHAIIVGYKTYQOIIQGIQOLVK	247
DB	169	TVEQAEQDIDIELTVAGT-KEAINWVESGAKESSEEDMGLALPGDAIKELVAPQEEIYA	227
QY	248	ETGVTKRTPQKLFTPSPEIIVKTYTKLAMERLY-----AVFTDYEHDKVSRDEAVNKIRLD	302
DB	228	AVGPKRMVDLL-----QVDADLKEEIFPAYNTMKTAVMT-----EKLAREVIDKVKYT	279
QY	303	TBEQLEKEFPF-ADPYEIIIESFNVA-----KEVPSIVLNEKYKCDGRDLTSLRNVSCV	357
DB	280	VKEVYAEKFSRHEEBAQILKEVQQAIEADLEKQVAVELLITDIKIRPDGRGLDIEIRHLSSEV	339

```

Qy 358 DMKTLHGALFORGOTVU/LCWTFPDSJESGKSDOVITANGKD--JNMLXHPPPV 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 SILPRHSGGLTTRGTOALSVCTI---APLGEHQIIDL-GVQDSKRRTIHHNPPQS 394
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 417 TNEIGVYTOGLNRRELGHGALAEKALPYIP--RPFPPTIRVTSHEVLSNGSSMASCG 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 VSGTGAGSGPRREIGHGALGERALAOIIPSEDPPTIIRVAVLESNGSSSASGSI CAG 454
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 475 SLALMDGVPISSAVAGVAGLVTTKDBKEIEDYRLITDIIGIEDYNGDMDFKAGTN 534
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 TALMDAGPIKAPVAGIAMGLVS-----DGENYITLTDIOGLBDHLDMPFKVAGTK 507
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 535 KGITLQADIKLPGPIPIKIWNKAIQOASVAKKEILOIMNKTIKSPRPSRKENGWVETVQ 594
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 DGITLQMDIKIOGITEOILTEALDOAKKAMEILBELVTTTIAAPREELSQVAKIEMIO 567
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 595 VPLSRKAKVSGPGVNLKKLQAEFGVTSOVDENFSGVAPTPSVMBHABDFTEICKOD 654
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 IKPAKIKDVIYKGGETINSIIDETGVKIIDODGQNVSTASSDAMIKKAIKIIIBELTK-- 625
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 655 QBOOLEFAGVYATITIELRDTGVMWKLXPNNMTAVLILHTQDSDNERLN 701
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 626 ---EVEVGVOYLAKVRIIEKCAFYNLIKDG--LIHISQLANRNV 668
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
 US-09-809-665A-171
 Sequence 171. Application US/09809665A
 Patent No. 6790950
 GENERAL INFORMATION:
 APPLICANT: Lowery B., David, et al.
 TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
 FILE REFERENCE: 28341/00435
 CURRENT APPLICATION NUMBER: US/09/809,665A
 CURRENT FILING DATE: 2001-03-15
 PRIOR APPLICATION NUMBER: 60/153,453
 PRIOR FILING DATE: 1999-09-10
 PRIOR APPLICATION NUMBER: 60/128,669
 PRIOR FILING DATE: 1999-04-09
 PRIOR APPLICATION NUMBER: 09/545,199
 PRIOR FILING DATE: 2000-04-06
 NUMBER OF SEQ ID NOS: 197
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 171
 LENGTH: 716
 TYPE: PRT
 ORGANISM: Pasteurella (Mannheimia) haemolytica
 US-09-809-665A-171

Query Match	31.7%;	Score 1126;	DB 4;	Length 716;
Best Local Similarity	37.1%;	Pred. No. 1,36-96;		
Matches	251;	Conservative 131;	Mismatches 255;	Indels 40; Gaps 12

QY	35	GNRKLEISSGCLAFAPDGSAAVVGSDTAAVMYTAWSKTKPSFSQ--FMPLVNDYRQCKLAAG	93
Db	11	GOHTVLTETGAIAGQATPAVAASNDTTFVTVAKKDVKEGQDFPLTVYQESTYTAAG	70
QY	94	RIPNTVYLREYVGTSDKELTSTRIDRSIRPLFPAGYFYDTQVLGNLAVDGNBBDVALI	153
Db	71	RIPGGFFRBERPREGETLLIRLLIDRPRLPFBPGFMEIGVIATVVSVNQDISPDVAM	130
QY	154	NGASVALSLSDI PNNNGPVGAVRIGIIDEEYVYNNPTKREMSSTLNLVYAGAPKSOIWMLE	213
Db	131	IGASAAALSLGVPNPGPFGAARVGFINDQFLANLPTTSQKISRDLVVSQGDKA-VLWVE	189
QY	214	ASAENIILQODFGAIIKGVKVTYQOIIQGIQOLVBEIVTXX--RTPOKLTFTSPSEIVKXT	270
Db	190	SEADILNBEQMLAAVNVGHEQQVIVENIKEFVEADGKPRWDVAPE---PNTDLINKV	245
QY	271	HKLAMERLYAVFTYENDKVSRLDAVANKIRLDTBEOLEKRRPEADRYIIISFPNVAKEV	330
Db	246	KALAEIRIGADYRIVE--KQVRYQIDAIKKEVIAQLAABEIVSEGTIIIDITALSESQI	303

SEQ ID NO 2259
LENGTH: 699
TYPE: PR
ORGANISM: M.cacarthalis
US-09-540-236-2259

Query Match 31.0%; Score 1103; DB 4; Length 699;
Best Local Similarity 38.1%; Pred. No. 1.8e-94;
Matches 260; Conservative 119; Mismatches 254; Indels 50; Gaps 14;

35 GNRKLEISGKLAFADGSAVVGSDTAVMATAVSKTPSPSQ-FMPLVVDYRQKAAAG 93
16 GNCQVLETRIAQAN-SVLVHMGVSLVAVVREAIAGQFFPLTVYQEKMAVG 74
94 RIFPNYARREVGTGDKELTSRIIDRSRRLFPKGYFIDYOVLCNLAVDGNEPDVLA 153
75 KIPGGYKRGRASEFETLSRLIDRPLRPLFPGYFNEIQVATVATVSSDKTOADIAAM 134
154 NGASVALSLSDIPNNGPVGAVRIGIIDEYVNPTRKEMSSSTNLVVGAPKSOIWMLE 213
135 IGASALAIAPAPNGPIGAAKAVFINGEYVLANPTLAEMKQSDLDLVAGT-KSAVLME 193
214 ASANILQODFCHAIKVGVKTYQOIIOGIQOLVKEGVTKRPOKFLTPSPBIYKTHKL 273
194 SEAKELSEDQMLGAVLXGHGQGVVINDINTFAQAVGNTK--QEFVAPA----- 240
274 AMELVAVFTDYENDKXSRDAVANKIRLDTBEOLKEKPEPA-----DPY-----EI 319
241 INEELNTQLKEQFTAKVSEAYTIR-VKODRYARLDELAAELVLAGETSDVDYAKVAOI 299
320 IESFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCVDMFKTLHSGALFORQTOVLCT 379
300 KEEFETLKRIVRNILNSGKPRIDGRDLETVALDIOGVLPYTHGSALEFTRGETOALV 359
380 VTFDSLESGIKSDVITAINGIKDNFMALHYEPFPYATNEIGKTYGTLNRRLEGALAEK 439
360 TLTGTLT---TRDVNLVDTLGTQDHFMLHPRPHYSVGTGREGARREIRIGRLARR 415
440 ALYVIR-RDFFTRIVTSVLENSGSSMAACGSLAMDGVPISSAVAGVAGLV 497
416 GVOAMLAERFPPTIIVSEITSSGSSMAVSCASLSLMDGVPLKAVAGIANGLV 475
498 TKTDPEKEIEDYLLTDILGIEDYNGMDPKIAGTNGGITALADIKLPGIPKIYMEA 557
476 K-----EGERFVLSLIDDEDLGMDPKVAGVNGITALLQMDIKIBITADIMEQA 528
558 IQQASVAKKEILQIMNKTISKPRASRKENGVEVTVQVPLSKRAKFGVGGYNLKKLOAE 617
529 LKQAHAGRHIILNMEVIATSRTEINAHAVNATITINPEKINDVIGKGAATIRLTED 588
618 TGVVISQVDETFEVPAPTSVMEHARDFTIEICKDQEOLEFGAVYTAITTEIRDTGV 677
589 TGAITIDIDDDGTITIFGTDKA---STRAATQI--EATIAVEVGTVEGVARIIVFGA 643
678 MVKLYPMNTAVLINTOLDNERL 700
644 FTITLPG-TDGLVHISQISDERV 665

RESULT 8
US-09-809-665A-109
Sequence 109, Application US/09809665A
Patent No. 6790950
GENERAL INFORMATION:
APPLICANT: Lowery E., David, et al.
TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
FILE REFERENCE: 28341/00435
CURRENT APPLICATION NUMBER: US/09/809,665A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: 60/153,453
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: 60/128,689
PRIOR FILING DATE: 1999-04-09

PRIOR APPLICATION NUMBER: 09/545,199
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 109
LENGTH: 684
TYPE: PR
ORGANISM: Pasteurella multocida
US-09-809-665A-109

Query Match 31.0%; Score 1101.5; DB 4; Length 684;
Best Local Similarity 37.8%; Pred. No. 2.3e-94;
Matches 246; Conservative 133; Mismatches 217; Indels 35; Gaps 13;

60 DTVMTAVSKTKPSQ-FMPLVVDYRQKAAAGRIPTYNLREVTSDEKILTSIID 118
6 DTVFVTVAAKKOVKESQDFFPLTVNQEITYAAGRIPGEPFKREGPSSGETLIAIID 65
119 RSIRPLFPAGYFYDTQVLCNLAVDGNEPDVLAINGASVALSLDIPNNGPVGAVRIGI 178
66 RPIRPLFPEGFYNEIQVATVSVNPOICPDVAMIGASALSLSGVFPNGPIGAARVGF 125
179 IDEGVNPTFKEMSSSTNLVVGAPKSOIWMLEASANILQODFCHAIKVGVKTYQOI 238
126 IDQFVLNPTMNEQKQRLDLVAGTDKA-VLWVSEADVLTEBOMLAADVFGHQOQOV 184
239 IOGIQOLVKEGVTKRPOKFLTPSPBIYKTHKLAMEBLVAVFTDYENDKXSRDEA 295
185 IDAKETTAAGKPRMDVAP---PNTALIEKVAIAARIGEARITE--KQAYEQ 238
296 VNKIRLDTBEOLKEKPEADPY---EIESFNVAKEVFRSIVLNEYKRCGRDLTSLRN 352
239 IDAIKADVIAQITAEVAGEGDISBGKIVDIFLAEQIVASRIIAGSPRIDGRTVTVRA 298
353 VSCGVDMFKTLHSGALFORQTOVLCTVTFDSLESGIKSD-QVITAINGIKDNFMALHYE 411
299 LDICTGVLPYTHGSALEFTRGETOALVATL-----GTERDAQIIDEITGERSHDFLHYN 353
412 PPTATNEIGKTYGTLNRRLEGALAEKALVPIYR-RDFFTRIVTSVLENSGSSMA 469
354 PPTVSGETMIGSPKREIRIGHGLARGVAAVMTLAEPPYVVRVSEITSSGSSMA 413
470 SAGGSALAMDGVPISSAVAGVAGLVTKTDPEKEIEDYRLTDLGIEDYNGMDPK 529
414 SVCGASALMDAGVPIAAVAGIANGLVK-----EDERFVLSLIDGSDHIGMDPK 466
530 IAGTNGITALQADIKLPGIPKIYMEALIQQASVAKKEILQIMNKTISKPRASRKENG 589
467 VAGTRGTVALQMDIKIEGITAIRIMQALNOAKSARLHIIGVMEQALPAPRADISDPAPR 526
590 VETVQVPLSKRAKFGVGGYNLKKLOEFGVTSQVDETFEVPAPTSVMEHARDITE 649
527 IYTMKIDPKIKIDVIGKGAATIRLTTEETSTSIDDDGVTKLAADVAG---NSAKVWAR 583
650 ICKDQEOLEFGAVYTAITTEIRDTGVMTKLYPMNTAVLINTOLDNERL 700
584 I--EDITAEVAGVAGVTKTRLADFGAFVSYGNKEG-LVHISQISDERV 631

RESULT 9
US-09-328-352-8112
Sequence 8112, Application US/09328352
Patent No. 6562858
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8112
LENGTH: 705

TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8112

Query Match 29.9%; Score 1065; DB 4; Length 705;
Best Local Similarity 36.8%; Pred. No. 6.7e-91;
Matches 250; Conservative 124; Mismatches 261; Indels 44; Gaps 14;

```
QY 35 GNRKLEISSGKLAFADGSAAVVGSDTAVMVTAVSKTPSPQPMPLVVDYRQKAAAGR 94
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 22 GQHVAVLETGRVAQAMTVLITMGVTVLVAVVAAPFAKAGQDPFLTVNVEQKQVAGR 81
QY 95 IPTNYLRREVSTQSKELITSRIRPLFPAGFYPTQVLCNLLAVDGVNEPVLAIAIN 154
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 82 IPGGYGRKGRSAEATLISRLIDRPIRPLFPBEGYNEIQVATVAVSDKTMEDIAAML 141
QY 155 GASVALISDIPMNGPVGAIVRIGIIDGEVYVNPTRKEMSSSTLNLVAGAPKQIWMLEA 214
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 142 GTSALALAGTPFGPFGAARVGLNGEYVLPNPFQMQSDLDLVVAGT-ESAVLAVES 200
QY 215 SAENILODFCHAIKVGKVTQOIIQGIQQLVKETGVTKTPQKLFPSPEIVKTKLA 274
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 201 EAKELSDQMGLAVLFGHDEMQIAQAINEPAAAAG--KPSDWVAPA-----HN-- 248
QY 275 MERLYAVFTDYEHDKVSR--DEAVNKIR--LDT--BEOLKEKFPADPYEIISEFNVA 327
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 249 -BEIRAKLKKEAFKAEISAVYTIKAKODRYALDLALHAFAVQFVPEEDVDGIDAEVDYLF 307
QY 328 KEV-FRST--VLNBYKRCQDGRDLSLRNVSCEVDMFETLHGSALFORQOTVLCVTFD 383
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 308 EDLKRIVRDNILSGKPRIDGRDITKTVBALVQGVLEFRAHGSALFTRGETQALVTTLG 367
QY 384 SLESKISDQVITAINGIKDKNFMILHYEPFPYATNIEIGKVTGLNRRELGHALAEKALYP 443
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 368 NTRDAL----MVDPLAGTKTDNFMILHYNFPAYSGEGRSGSPRRREIGHRLARRGQA 423
QY 444 VTPR--DPPFTIRVTSVLESGSSSMASACGSLAMDGVPISSAVAGVATGLVTKTD 501
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 424 VLPADRFPPYIRIVSDITBESGSSSMASVCGASLSIMDAEVPKAPAYAGIMGLVYK-- 480
QY 502 PEKGEIEDYRLTLTILGIEDYNGMDPKIAGTNKGTALQADIKLPGPIKIYMEALIOQA 561
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 481 ----EGERFVALSLILGDEDEHLGMDRFKVASANGITRLQMDIKIETIGTIEIMVALNOA 536
QY 562 SVAKKEILLQNMKTIKSPRASKENGVPVETVPLSKRAKFPVPGYNLKKLQAEITGV 621
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 537 FAGMHILINEMKXVTSRAREPISMAHPFEVITINPKIRIVIGKGAITRQIETEKAA 596
QY 622 ISQVDEETFSVFAPTPSVMEHARDFTIEICKDOEQLEFGAVYTAITTEIRDTGVNKL 681
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 597 IDIDNGTVRVFGETKAAAKAIAIKQAI-----TAEVEPKIYDGKIVRIVEFGAFVNI 651
QY 682 YPNMTAVLHNTQLDNERL 700
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 652 MFG-TDGLHLISQLSNERI 669
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 10
US-09-583-110-2924
Sequence 2924, Application US/09583110
Patent No. 669703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PATHO0-07A
CURRENT APPLICATION NUMBER: US/09/583,110
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
```

PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 2924
LENGTH: 737
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-2924

Query Match 29.6%; Score 1053; DB 4; Length 737;
Best Local Similarity 34.4%; Pred. No. 9.7e-90;
Matches 233; Conservative 150; Mismatches 256; Indels 38; Gaps 9;

```
QY 37 RKLFISSGKLARADGSAAVVGSDTAVMVTAVSKTPSPQPMPLVVDYRQKAAAGRIP 96
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 13 RELIVERGVAKAQSAAVVRVYESTVLTAAVSKKQATDFFPLQVNEKKYAAAGKFP 72
QY 97 TNYLRREVSTQSKELITSRIRPLFPAGFYPTQVLCNLLAVDGVNEPVLAIAINGA 156
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 73 GGFMRKGRPSTAITLRLIDRPIRPLFMRGFRNEVQVINTVYLSIDENASAMAMAFGS 132
QY 157 SVALISDIPMNGPVGAIVRIGIIDGEVYVNPTRKEMSSSTLNLVAGAPKQIWMLEASA 216
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 133 SLALISDIPDPPIAGVQGVYDQGIINPSQEQAEQSLLELTAVGT-KHAVNVESSGA 191
QY 217 ENILODFCHAIKVGKVTQOIIQGIQQLVKETGVTKTPQKLFPSPEIVKTKLAME 276
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 192 KEISEEIMELALKHGAHVALLAFOBEIYAAVAKK-----AEVELLHVDAELQAE 243
QY 277 RLVAVFPTYE----HDKVRDEAVNKKIRLDTBEOLKEKFPADPY-----EIISEFNVA 326
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 244 TIAAASDLQAVQVEKLAKEALATQVAVYVEKADHEPDRIMRVAELIEQM 303
QY 327 AKVFRSIVLNBYKRCQDGRDLSLRNVSCEVDMFETLHGSALFORQOTVLCVTFDLSL 386
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 304 EHAVRRLITFEDKVRPDGRKRVDEIRPLDAVVDLPVHSGSLFTFRQOTQALSVTLT---- 359
QY 387 SGKISDQVITAINGIKDKNFMILHYEPFPYATNIEIGKVTGLNRRELGHALAEKALYP 446
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 360 APNGETQIIDGLDEPKKRMTHNPNQYSVGETRGAGARREIGHGALGEEALQVLP 419
QY 447 --DPPFTIRVTSVLESGSSSMASACGSLAMDGVPISSAVAGVATGLVTKTDPEK 504
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 420 SLEFPFALRLVAEVLNESGSSQASICAGTLLMAGVPIKAPVAGIAMGLIS----- 473
QY 505 GEIEDYRLTLTILGIEDYNGMDPKIAGTNKGTALQADIKLPGPIKIYMEALIOQASVA 564
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 474 -DGNNTVTLVDIGLEBDFMDPKVAGTRDITLQMDIKIQTAEILTEBALQAKKA 532
QY 565 KKEILLQNMKTIKSPRASKENGVPVETVPLSKRAKFPVPGYNLKKLQAEITGVISQ 624
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 533 RFEILDVIEATTVIEVRLLELAPTPAKIDTIKIDVDKIKIYGGKGETTDKIETAEVAKIDI 592
QY 625 VDEETFSVFAPTPSVMEHARDFTIEICKDOEQLEFGAVYTAITTEIRDTGVNKL 684
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 593 DEGNVSIYSSDDAIDRAKEITAGLVREAKVDE-----YRAKAVRIEKGAFVLLF-D 646
QY 685 MTAVLHNTQLDNERLN 701
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 647 KTDALVHISEMAMWTRTN 663
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 11
US-09-252-991A-27283
Sequence 27283, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
```

PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 21283
 LENGTH: 697
 TYPE: PR
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27283

Query Match 27.5%; Score 976.5; DB 4; Length 697;
 Best Local Similarity 35.6%; Pred. No. 1.4e-82;
 Matches 228; Conservative 130; Mismatches 230; Indels 53; Gaps 15;

86 ROKAAAGRTPTNLR-----EVTSDKEI-----TSRIIRSRIRPLPAGFY 131
 45 RPKAATSRRCPTTRKPTPOAVSPAVSSVADVLPRTTSLRILDIRPLPPEGGMN 104
 132 DTQVLCNLADVNEPDLVLAINGASVALSLSDIPMNPVGAVRIGI-IDGEVYVNPTRK 190
 105 EVQVVCIVSTNKSDDPDIAAMIGTSALASIGIPFAGPICAARVGFHPELGIYLNPTYE 164
 191 EMSSTLNLVAVAPKSOIWMLEASAEENILQDFCHAIKGVKXTQOIIQGIQQLVKEG 250
 165 QLOSSSLDMVAVGT-EDAVLWVESEADELTEDQMLGAVLFAHDEFQAVIRAVKELAAEAG 223
 251 VTK---RTPOKLTFPSPETVYKTHKLAMERLYAVFTVDYEHDKV-----RDEAVNKIRLD 302
 224 KPAWDMKAPAE-NVLVNAIETAEIGEAISOAVTI--TIKDRYNRGLERQOAVALPAGE 280
 303 TEOLOKEFPADYEIIESFNVAKEVFRSIVLNEYKRCGRDLSLRNVSCVDMFKT 362
 281 EE----GKFPAS---EYVDVFGLEIRYVRENINVKGRIDGRDTRVRIEIVGLVK 333
 363 LHGSALFORQGTQVLCVTFPDSLESGLKSDQVITAINGIKDKNFMALHYEPPPVATNEIGK 422
 334 THGSALFTRGETQALVAVATLGT---ARDAQLDLTEGERKDAFMALHYNFPFVGEBCGR 389
 423 VTGNRELGHALAEKALYVPIR--DFPTIRVTSVLESNGSSSMAACGSLALMD 480
 390 MGSGRREIGHGRARGVAMLPDDEFPYTIIRVSEITSSSSSMAVSCGASLALMD 449
 481 SGVPISSAVAGVAGLVTKTDPKGEIEDYRLLDLIGIDYNGMDPKIAGTKNGITATL 540
 450 AGVAVKAPVAGIANGLVK-----EGEKAFLVDLIDGEDHLDMDPKVAGTDKGTATL 502
 541 QADIKLPQIPKIVMEALQOASVAKKEILQIMNTISKPRARKENGVEVETVQVPLSKR 600
 503 QMDIKINGITBEIMEIALGQALBARLNILGQMOVIAKPRABLESMAVPTMLQMKIDSDKI 562
 601 AKFVPGGNLUKKLOAETGVTSIOVDETFSVFAPTSVMEHARDFTTEICKDOEOOLE 660
 563 RDVIGKGAIRIGCEETKASIDIEDDGSVKIYGTEKMAEAALKIRVALI-----TAAE 617
 661 FGAVYATITETIRDTGVAVVKLYPMNTAVLHNTOLDNERLN 701
 618 IGTIVGKERTIVDFGAFLNLPKQDGLVAHSIQISDKRID 657

RESULT 12
 US-09-198-452A-1073
 Sequence 1073, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Glifalab, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 TITLE OF INVENTION: and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 PRIOR FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 1073

LENGTH: 568
 TYPE: PR
 ORGANISM: Chlamydia pneumoniae
 FEATURE:
 NAME/KEY: SITE
 LOCATION: 1...568
 OTHER INFORMATION: Xaa=unknown or other
 US-09-198-452A-1073

Query Match 24.9%; Score 884; DB 4; Length 568;
 Best Local Similarity 36.9%; Pred. No. 4.9e-74;
 Matches 205; Conservative 112; Mismatches 198; Indels 40; Gaps 13;

157 SVALSIDIIPMNPVGAVRIGIIDGEVYVNPTRKEMSSSTLNLVAVAPKSOIWMLEASA 216
 8 SALAISIDIPQSNVAVAGRICIDNOMVINPTTELASSLTDVLVAGT-ENAILMIEGHC 66
 217 ENILQDFCHAIKGVKXTQOIIQGIQQLVKEGTVKTRPOKLTSPSEIVKXTTKLAME 276
 67 DFTPEQVLDALIEGHGHIYVICRQLQMOEVBGSKNLSAVVPLPA-EVLTAVKECAQD 125
 277 RLAVFTVDYEHDKVSRDEAVNKIRL-DTEBOLKEFPADYEIIESFNVA-----KE 329
 126 KFTLEFN-----IKDKVHAATFAHEIENITLKEQLQREDD-DLFSSFNITKAACKTLKSD 177
 330 VFRSIVLNEYKRCGRDLSLRNVSCVDMFKTLHGSALFORQGTQVLCVTFPDSLES 387
 178 TMRALINDREIRADGRSLTVRPTTITSYLPRTHSGCLTRGETQVLAICTLQSEAMA- 236
 388 GIKSDQVITAINGIKDKNFMALHYEPPPVATNEIGKTVGNRELGHALAEKALYVPIR- 446
 237 -----QREYDLNGEGLSKFYLYQFFPPFVGEVGRIGSPRRREIGHGKLEKALSHALPD 291
 447 -RDPEPTIRVTSVLESNGSSSMAACGSLALMDGCVPISSAVAGVAGLVTKTDEKG 505
 292 SATPPTIRIESNTTESNGSSSMAVSCGGLALMDAVALPSSPAGIAGMLIL---DDQG 348
 506 EIEDYRLTDLIGIDYNGMDPKIAGTKNGITATLQADIKLPQIPKIVMEALQOASVAK 565
 349 AI-----ILSDISGLEHLDGMDPKIAGSGKITRPFQMDIVKEGITRPMKALSOAQGC 404
 405 NDILINMNAISAPKADLSQAPRIETMQIKPTKIASIVIPGQKQRIQIIEETVQIDVN 464
 626 DEETFSVFAPTPSVMEHARDFTTEICKDOEOLEFNAVTTATITETIRDTGVAVVKLYPM 685
 465 DLGAVSISASSASAINKAKEIIEGLV-----GEVEVGKTYRGRTSVVAFGAFVEVLPGK 519
 686 TAVLHNTOLDNERL 700
 520 EG-LCHISEGSRORI 533

RESULT 13
 US-09-270-767-44788
 Sequence 44788, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:
 APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 PRIOR FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 44788
 LENGTH: 332
 TYPE: PR
 ORGANISM: Drosophila melanogaster
 OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-44788

```
Query Match 24.2%; Score 860.5; DB 4; Length 332;
Best Local Similarity 52.4%; Pred. No. 3.1e-72;
Matches 174; Conservative 48; Mismatches 107; Indels 3; Gaps 2;

QY 211 MLESAENIILOODCHAIKVCVXTQOIQIIOQLVETGYTKRTPQKLFPSPEIYKYT 270
DB 1 MLEKGNVAVLOODLLKAIKQGTREAOPTIHEIERIQKAGKQKEVEVAAEVDLGA 60
QY 271 HKLAMELYAVFTYEHDKVSRDAVNKIRLDTSEQLKEKPEADPYEIIISFNVVAKV 330
DB 61 RACKCGEKSSRRQRPRAKISRANAVNEVRSNVLDKWSPPDEPESLITQFQGTERTI 120
QY 331 FRSLVLEKRCQDGRDLTSLRNVSCEDVMFKTLHGSALEFGQGTQVLTCTVFDLSG 390
DB 121 FRELIFERGLRCDGRDYQLTNIISQYDMYKPLHGSALEFGQGTQVCTVSLDSQES 180
QY 391 SDQV--ITAINGIDKXKPMFLHYEPFPYATNIGVTLGNRELHGALEAKLYPVIP 448
DB 181 LDSLAALDSXXXXPSPNFMFLHYEPFPYATGSGRIGPGRREMGHGALEKSLPLT 240
QY 449 PPTIRVTSVLESGSSSMASACGSLALMDSGVPISSAVAGVAILGVTK--TDPKEGEI 507
DB 241 YPFTVRLTSVLESGSSSMASVCGSLALMDAGVPVSAAPAGVAILGLVTKFENDTK 300
QY 508 EDYRLTDILIGIEDYNGDMDEKIACTNKGITA 539
DB 301 QDYRLTDILIGIEDYNGDMDEKIACTNKGITA 332

RESULT 14
US-09-809-665A-44
; Sequence 44, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-809-665A-44

Query Match 20.0%; Score 710; DB 4; Length 488;
Best Local Similarity 36.4%; Pred. No. 8.7e-58;
Matches 162; Conservative 87; Mismatches 170; Indels 26; Gaps 9;

QY 262 PSPPIVKTHTKLAMELYAVFTYEHDKVSRDAVNKIRLDTSEQLKEKPEADPY---E 318
DB 11 PNTLLIEVKVKAIAENARIGEARITENKHYM--NKDIAKADVIAQITRAVAEGEDISE 68
QY 319 IIESFNVVAKVFRSIVLEKRCQDGRDLTSLRNVSCEDVMFKTLHGSALEFGQGTQV 378
DB 69 IVDLFTALSESQIVASRIIAGEPRIDGRVDVLRALDICTGLPRTHSGAIIITRGETQ 128
QY 379 TTFPDSLESIGKSD-QVITAINGIDKXKPMFLHYEPFPYATNIGVTLGNRELHGA 437
DB 129 VATL-----GTERDAQIITDELGTGRSDHFLFHYNFPYVSVGETGMISGPKREI 183
QY 438 EKALYPIV--RDPFTIRVTSVLESGSSSMASACGSLALMDSGVPISSAVAGVAILG 495
DB 184 KRGVAAVMTLAEFPYVAVSVETTESGSSSMASVCGSLALMDAGVPVIAAAGVAILG 243
```

```
QY 496 LVTKTDPKEGEIEDYRLTDILIGIEDYNGDMDEKIACTNKGITALQADILPGIPIKI 555
DB 244 LVK-----EDKEFVLSIDILDEDHLDGMDPKVAGTRGVATLQMDIDIEGITA 296
QY 556 EAIQOASVAKKELIQTQNMKTIISKPRASRKENGVEVTVQVPLSKRAKFPVGGYNI 615
DB 297 IALNOASALHLILGWMEQAIAPBRADISDPARITVMTKIDPKKIKDVIKGGATTA 356
QY 616 AETGVTSIQVEERTFSVAFPTPSVMEHARDPITEICKDDEQQLERGAVYATIT 675
DB 357 EBTGTSIDIDDDGTVXIAAVDG--NSAKEVMARI--EDITAVEAGAVYKGVTRLADF 411
QY 676 GVMVKLYPNMTAVLLNTOQIDNERL 700
DB 412 GAFVSIQNKEG-LVHISQIAEERV 435

RESULT 15
US-09-809-665A-155
; Sequence 155, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Actinobacillus pleuropneumoniae
US-09-809-665A-155

Query Match 17.9%; Score 636; DB 4; Length 358;
Best Local Similarity 38.3%; Pred. No. 4.7e-51;
Matches 143; Conservative 67; Mismatches 139; Indels 24; Gaps 7;

QY 242 IOOLVKETGVTK--RTPQKLFPSPEIYKYTKLAMELYAVFTYEHDKVSRDAVNK 298
DB 2 IKEFVKEAGKPRMDWAPF-----PNTALINQVAKALAEARIGDAVRLITE--KQ 55
QY 299 IRLDTSEQLKEKPEADPYEIIISFNVVAKVFRSIVLEKRCQDGRDLTSLRNVSC 358
DB 56 IKADVIAQLTAQOQETVSEGAIIIDITALBESSIVGRRIAGEPRIDGRVDVLRALD 115
QY 359 MFKTLHGSALEFGQGTQVLTCTVFDLSGIGKSD-QVITAINGIDKXKPMFLHYEP 417
DB 116 VLPRTHSGAIIITRGETQALAVATL-----GTERDAQIIVDELTEKSDRFLF 170
QY 418 NEIGKVTGLMRRRLHGALEAKLYPIV--RDPFTIRVTSVLESGSSSMASACG 475
DB 171 GEFGRGSPKREKRIHGRLLKRGVAVLMPAEFFPVYRVSVITTSNGSSMASVCG 250
QY 476 LALMDSGVPISSAVAGVAILGVTKDPKEGEIEDYRLTDILIGIEDYNGDMDEK 535
DB 231 LALMDAGVPVIAAAGVAILGVLK-----EKEKFVLSIDILDEDHLDGMDPKV 283
QY 536 GITALQADILPGIPIKIYMEALIQOASVAKKELIQTQNMKTIISKPRASRKENG 595
DB 284 GTVALQMDIDIEGITEIMQIALNOAKGARMHLSVMEQAIAPBRADISDPARIT 343
QY 596 PLSKRAKFPVGG 608
DB 344 DPKKIDVIGKGG 356
```


Search completed: January 28, 2005, 19:36:29
Job time : 43 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:35:48 ; Search time 147 Seconds
(without alignments)
1732.714 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGFLLPRRDALVQLQVRA.....TAVVLTHTQDNERLTLLP. 705

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications MA:
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3557	100.0	705	US-09-907-907A-42	Sequence 42, Appl
2	3557	100.0	705	US-09-907-907A-44	Sequence 44, Appl
3	2285	64.2	504	US-10-408-765A-1791	Sequence 1791, Ap
4	1267	35.6	1526	US-10-437-963-117217	Sequence 117217,
5	1204.5	33.9	703	US-10-282-122A-51965	Sequence 51965, A
6	1191.5	33.5	696	US-09-815-242-5443	Sequence 5443, Ap
7	1191.5	33.5	698	US-09-815-242-12325	Sequence 12325, A
8	1190.5	33.5	698	US-10-282-122A-44413	Sequence 44413, A
9	1165.5	32.8	701	US-10-282-122A-70948	Sequence 70948, A
10	1161.5	32.7	692	US-10-282-122A-71588	Sequence 71588, A
11	1161	32.6	711	US-10-282-122A-75529	Sequence 75529, A
12	1158	32.6	721	US-09-815-242-13768	Sequence 13768, A
13	1157	32.5	729	US-10-282-122A-61234	Sequence 61234, A

14	1155	32.5	703	15	US-10-282-122A-53063	Sequence 53063, A
15	1155	32.5	734	9	US-09-815-242-10312	Sequence 10312, A
16	1155	32.5	734	15	US-10-282-122A-56702	Sequence 56702, A
17	1147	32.2	711	15	US-10-282-122A-59492	Sequence 59492, A
18	1146.5	32.2	713	15	US-10-282-122A-47769	Sequence 47769, A
19	1146	32.2	705	15	US-10-282-122A-77935	Sequence 77935, A
20	1140	32.0	709	15	US-10-282-122A-68773	Sequence 68773, A
21	1135.5	31.9	719	15	US-10-282-122A-51291	Sequence 51291, A
22	1135	31.9	712	15	US-10-282-122A-56195	Sequence 56195, A
23	1134	31.9	709	9	US-09-815-242-10989	Sequence 10989, A
24	1134	31.9	709	15	US-10-282-122A-58109	Sequence 58109, A
25	1132	31.8	705	10	US-09-907-907A-43	Sequence 43, Appl
26	1131.5	31.8	771	15	US-10-282-122A-48837	Sequence 48837, A
27	1128.5	31.7	704	9	US-09-815-242-10882	Sequence 10882, A
28	1128.5	31.7	704	15	US-10-282-122A-42421	Sequence 42421, A
29	1127.5	31.7	763	15	US-10-282-122A-50508	Sequence 50508, A
30	1127	31.7	702	9	US-09-815-242-4975	Sequence 4975, Ap
31	1127	31.7	723	15	US-10-282-122A-60483	Sequence 60483, A
32	1126	31.7	716	11	US-09-809-665A-171	Sequence 171, Appl
33	1126	31.7	716	17	US-10-854-299-171	Sequence 171, Appl
34	1125.5	31.6	714	15	US-10-282-122A-67188	Sequence 67188, A
35	1121	31.5	709	15	US-10-282-122A-77113	Sequence 77113, A
36	1117.5	31.4	714	15	US-10-282-122A-49590	Sequence 49590, A
37	1117	31.4	707	15	US-10-282-122A-65516	Sequence 65516, A
38	1114.5	31.3	694	15	US-10-282-122A-54319	Sequence 54319, A
39	1112	31.3	712	15	US-10-282-122A-45343	Sequence 45343, A
40	1111	31.2	706	15	US-10-282-122A-65636	Sequence 65636, A
41	1111	31.2	706	15	US-10-275-026A-36	Sequence 36, Appl
42	1109	31.2	732	15	US-10-282-122A-57601	Sequence 57601, A
43	1107	31.1	706	15	US-10-282-122A-52565	Sequence 52565, A
44	1103	31.0	695	15	US-10-282-122A-62955	Sequence 62955, A
45	1101.5	31.0	684	11	US-09-809-665A-109	Sequence 109, Appl

ALIGNMENTS

RESULT 1
US-09-907-907A-42
Sequence 42, Application US/09907907A
Publication No. US20030099660A1
GENERAL INFORMATION:
APPLICANT: Fisher, Paul B.
TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A
FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
CURRENT FILING DATE: 2001-07-16
PRIORITY APPLICATION NUMBER: US/09/907,907A
PRIOR FILING DATE: 1999-02-02
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 42
LENGTH: 705
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-907A-42

Query Match 100.0%; Score 3557; DB 10; Length 705;
Best Local Similarity 100.0%; Pred. No: 7.8e-270;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGFLLPRRDALVQLQVRAVVDGNRLTETSSGKLAFAGSNAVVOGSD 60
DB 1 DGFLLPRRDALVQLQVRAVVDGNRLTETSSGKLAFAGSNAVVOGSD 60
QY 61 TAVVTAASKTPSPQSPQMPVAVYRQKAAAGRIPTVYLRREVGTSDKEITLTSRIIDRS 120
DB 61 TAVVTAASKTPSPQSPQMPVAVYRQKAAAGRIPTVYLRREVGTSDKEITLTSRIIDRS 120
QY 121 IRPLPAGFYDYOVLGNLAVDGVNEEDVLAINGASVALSLSDIPMNGVGAVRIGIID 180
DB 121 IRPLPAGFYDYOVLGNLAVDGVNEEDVLAINGASVALSLSDIPMNGVGAVRIGIID 180

```

Db      121  IRPLFPAGFYFDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNGPVGAVRIGITID 180
Qy      181  GEYVNPTRKEMSSSTLNLVVGAPKSOIWMLEASAEINILQODFCHAIKVGKTYTOOIIQ 240
Db      181  GEYVNPTRKEMSSSTLNLVVGAPKSOIWMLEASAEINILQODFCHAIKVGKTYTOOIIQ 240
Qy      241  GIQQLVKETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db      241  GIQQLVKETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Qy      301  LDTEBOLKEKFPPEADPYEIIIESFNVAKEVRSIVLNEYRCODRDLTSLRNVSCEYDMF 360
Db      301  LDTEBOLKEKFPPEADPYEIIIESFNVAKEVRSIVLNEYRCODRDLTSLRNVSCEYDMF 360
Qy      361  KTLHGSALFORGOQVLCYVTFPDSLESGIKSDQVITAINGIKDNFMFLHYEPPYATNEI 420
Db      361  KTLHGSALFORGOQVLCYVTFPDSLESGIKSDQVITAINGIKDNFMFLHYEPPYATNEI 420
Qy      421  GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLSENGSSMASACGSLALMD 480
Db      421  GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLSENGSSMASACGSLALMD 480
Qy      481  SGVPISSAVAGVAILGYTKTDPKEGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540
Db      481  SGVPISSAVAGVAILGYTKTDPKEGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540
Qy      541  QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600
Db      541  QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600
Qy      601  AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660
Db      601  AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660
Qy      661  FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705
Db      661  FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705

RESULT 2
US-09-907a-44
; Sequence 44, Application US/09907907A
; Publication No. US20030099660A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Leszczyniecka, Magdalena
; TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A
; TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
; FILE REFERENCE: A34584-A-PC-T-USA (070050.1664)
; CURRENT APPLICATION NUMBER: US/09/907, 907A
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 09/243,277
; PRIOR FILING DATE: 1999-02-02
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-907a-44

```

```

Query Match      100.0%; Score 3557; DB 10; Length 705;
Best Local Similarity 100.0%; Pred. No. 7.8e-270;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  DGPELPRRRALTOLOVRALMSAGSRAVAVDIGNRKLETSSGKLARFADGSAVVOGSD 60
Db      1  DGPELPRRRALTOLOVRALMSAGSRAVAVDIGNRKLETSSGKLARFADGSAVVOGSD 60
Qy      61  TAVVVTAVSKTKPSQFMPLVVDYRQKAAAGRIPTVYLAREVGTSDKEILTIRIIDS 120
Db      61  TAVVVTAVSKTKPSQFMPLVVDYRQKAAAGRIPTVYLAREVGTSDKEILTIRIIDS 120

```

```

Qy      121  IRPLFPAGFYFDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNGPVGAVRIGITID 180
Db      121  IRPLFPAGFYFDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNGPVGAVRIGITID 180
Qy      181  GEYVNPTRKEMSSSTLNLVVGAPKSOIWMLEASAEINILQODFCHAIKVGKTYTOOIIQ 240
Db      181  GEYVNPTRKEMSSSTLNLVVGAPKSOIWMLEASAEINILQODFCHAIKVGKTYTOOIIQ 240
Qy      241  GIQQLVKETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db      241  GIQQLVKETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Qy      301  LDTEBOLKEKFPPEADPYEIIIESFNVAKEVRSIVLNEYRCODRDLTSLRNVSCEYDMF 360
Db      301  LDTEBOLKEKFPPEADPYEIIIESFNVAKEVRSIVLNEYRCODRDLTSLRNVSCEYDMF 360
Qy      361  KTLHGSALFORGOQVLCYVTFPDSLESGIKSDQVITAINGIKDNFMFLHYEPPYATNEI 420
Db      361  KTLHGSALFORGOQVLCYVTFPDSLESGIKSDQVITAINGIKDNFMFLHYEPPYATNEI 420
Qy      421  GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLSENGSSMASACGSLALMD 480
Db      421  GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLSENGSSMASACGSLALMD 480
Qy      481  SGVPISSAVAGVAILGYTKTDPKEGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540
Db      481  SGVPISSAVAGVAILGYTKTDPKEGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540
Qy      541  QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600
Db      541  QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600
Qy      601  AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660
Db      601  AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660
Qy      661  FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705
Db      661  FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705

```

```

RESULT 3
US-10-408-765A-1791
; Sequence 1791, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Wainock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408, 765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1791
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1791

```

```

Query Match      64.2%; Score 2285; DB 16; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.8e-170;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  DGPELPRRRALTOLOVRALMSAGSRAVAVDIGNRKLETSSGKLARFADGSAVVOGSD 60
Db      18  DGPELPRRRALTOLOVRALMSAGSRAVAVDIGNRKLETSSGKLARFADGSAVVOGSD 77

```

```

Qy 61 TAVVTAVSKTSPSPQEMPLVVDYRQKAAAGRIPTYLRREVSTSDKEILTSRIIDS 120
Db 78 TAVVTAVSKTSPSPQEMPLVVDYRQKAAAGRIPTYLRREVSTSDKEILTSRIIDS 137
Qy 121 IRPLPAGYFYDTQVLCNLAVDGNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
Db 138 IRPLPAGYFYDTQVLCNLAVDGNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
Qy 181 GEYVNPFRKMSSTLNLVVAAGAPKSGIWMLEASAEIILLODFCHAIKVGKTYQOIIQ 240
Db 198 GEYVNPFRKMSSTLNLVVAAGAPKSGIWMLEASAEIILLODFCHAIKVGKTYQOIIQ 257
Qy 241 GIQOLVETGYTKTPQKLTFTSPPEIVYTKHKLAMERLYAVFTDYEHKVSDEAVNKIR 300
Db 258 GIQOLVETGYTKTPQKLTFTSPPEIVYTKHKLAMERLYAVFTDYEHKVSDEAVNKIR 317
Qy 301 LDTEEQLEKEKPEADPYEIIESFNVAKEVRSIVLNEYKSCDGRDLSLRNVSCEVDMF 360
Db 318 LDTEEQLEKEKPEADPYEIIESFNVAKEVRSIVLNEYKSCDGRDLSLRNVSCEVDMF 377
Qy 361 KTLHGSALFQRCQOVLCCTVTPDSLESGIKSDOVITAINIKDKPMFLHYEPYATNEI 420
Db 378 KTLHGSALFQRCQOVLCCTVTPDSLESGIKSDOVITAINIKDKPMFLHYEPYATNEI 437
Qy 421 GKVTGLNRRELGHGALAEKALYPVIRPDRP 450
Db 438 GKVTGLNRRELGHGALAEKALYPVIRPDRP 467

```

RESULT 4

```

US-10-437-963-117217
; Sequence 117217, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrei A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 117217
; LENGTH: 1526
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(1526)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20643C.1.dep
US-10-437-963-117217

```

```

Query March 35.6%; Score 1267; DB 16; Length 1526;
Best Local Similarity 38.4%; Pred. No. 1,1e-89;
Matches 275; Conservative 137; Mismatches 240; Indels 64; Gaps 11;
Qy 33 DLGNRLKLEISSGKLARFADGSAAVQSGDTAVVNT-AVSKTKSPSPQEMPLVVDYRQKAAA 91
Db 66 EIGGRVVISFETGKARFANGSVISMDDTHLSTAAKSSRPVDFPLVTDYGEKQYA 125
Qy 92 AGRIPNTYLRREVSTSDKEILTSRIIDSIRPLPAGYFYDTQVLCNLAVDGNEBPVL 151
Db 126 QGVIPFTYMRREGAPKEREILLGRITIDRPIRLPFRPGYUHEQVIMVNVISDGGQDDPVM 185

```

```

Qy 152 AINGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPFRKMSSTLNLVVAAGAPKSGIWM 211
Db 186 AANSAAALMLSDIPMNGPVGIVIRGIDGNFVLPNTVDELGLSDNLVVA-CSRDKTL 244
Qy 212 LEASAEIILLODFCHAIKVGKTYQOIIQGIQOLVKEGYTKTPQKLTFTSPPEIVYTKH 271
Db 245 IDVAREITERDLOAGMKLAAEAHVCKINPOLRLAKRAG-KKKEKYKISLSDSVEYKIR 303
Qy 272 KLAMERLYAVFTDYEHKVSDEAVNKIRLDTEBOLEKEKPEADPYEIIESFNVAKEV 331
Db 304 TLSAPLEEVFTDSTYGKFERGELBENTTOSVAKLECEDESDSLFKHADVTRKQVI 363
Qy 332 RSVLNEYKSCDGRDLSLRNVSCEVDMFKTLHGSALFQRCQOVLCCTVTPDSLESGIKS 391
Db 364 RKRILIEKGLRVDGQDLDEVRPLCYESSYPIIHGSALFQRCQOVLCCTVTL-----GAPG 418
Qy 392 D-QVITAINIKDKPMFLHYEPYATNEIETKVTGLNRRELGH----- 433
Db 419 DAORLDSIVGPPTKRFMLHVSFPFESINEVAKRGLNRREVGHVYRNKPIRLGDELRYWI 478
Qy 434 -----GALAEKALYPVIRPDRP----- 465
Db 479 EHKALDPTGELLNRLYFWEKQWLSAGTLAEKALALVLPBGCEPFTYRANSEVMASSGS 538
Qy 466 SSMAACGSGSLALNDSGVPISSAVAGVAILGVTKTDPKEGIEBRYRLTDILIEDYNGD 525
Db 539 TSMAVCGSGMALNDAGIPVAEHVAGVSGLVSEVDQGTGDISSYRITDILGEDHLDG 598
Qy 526 MFKIAGTNKGITLQADIKLPGIPIKIWEAIOASVAKKEIIOINKTKISKRPAKKE 585
Db 599 MFKIAGTRRGITAIQIDIKPAGIPLDICELEPARKARQIILDRMOQESSRAAFNDG 658
Qy 586 NGPVETQVPLSRKAFVVGCGYNLKLQAEFTVTSIQVDETFVSVAFPVSHHEARD 645
Db 659 SSPRLATISSSSDLRKLL-----FHRKKIIEGTARVS-VSDGVTTVYAKTOPIDKALE 713
Qy 646 FITEICKDQEOLEFGAVYATITTEIRDTGVWKLXPNMTAVLHNTQDNERLN 701
Db 714 KVEFLV-----GREIEVGRTYGVVSIKEYGAFVFNCGGCG-LHHSLSHDKVS 764

```

RESULT 5

```

US-10-282-122A-51965
; Sequence 51965, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foretych, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

```

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 51965
;; LENGTH: 703
;; TYPE: PR1
;; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51965

Query Match 33.9%; Score 1204.5; DB 15; Length 703;
Best Local Similarity 39.6%; Pred. No. 2.6e-85;
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

QY 33 DLGRKLEISSGKLARFADGSAVVOGPTAVMTAVSTKSPSPS-QEMLVVDYRORAKAA 91
DB 8 DIARGLKVECKGTGMLSNCFISYGDVVMVNVNASEKREGIDFPLSIEBEROYS 67
QY 92 AGRIPVYLRREVSTDEKELTSRIIDRSIRPLPAGYFYDTQVLCNLAVDGVNEPDVL 151
DB 68 VGKIPGVKRGKRGSEKSTIHARALDRPLRPFKGYRNDVQVCTWVSVEQNLPEIL 127
QY 152 AINASVALSLSDIPMNGPVCAVRIGITDGYVNPTRKEMSSSTLNLVAGAPKSQIVM 211
DB 128 AMNAGSVALCTSDIPFTTPVAVTSVGCIDGKFLNPLTEBERKSSLDLTVC-ATNERVYM 186
QY 212 LEASAEINLQDDFCHAIKVGVKTYQOIIQGLQOLYKENVTRKTPOKLFTSPSPILVYKTH 271
DB 187 LEAGADEIPEDMLAALIDPFGNACDIAFOEKKMEKGEKATPE-LYHKEKIEKDV 245
QY 272 KLAMERLYAVFTDYEHDKVSRDEAVNKRILDTESQLEKEPPEADPYEIESFNVAKEVF 331
DB 246 EFAPESIKELM-YITDRDENLRLREIKESINBPKEKYPD-DGADIDEVYVTLQKKV 302
QY 332 RSIYVNEYKCDGDLTSLRNVSCEDVMFKTLHGSALFQRCQTOVLCTVTDSLSGKIS 391
DB 303 RNMLLKEHRBDGRFDEIRIPISCDVDLPRTHSGGLFTRGLTOVMVTTLV---GPICD 358
QY 392 DQVTAINGIKDKNPMLEHFFPPATNEIGKVTGILNRELGHALAELKALVPVP-RDF 449
DB 359 AQVIDGLGVESKRYMHYNFPYISGEVKPLRGPNRKEIGHALAEKALVPLIPSEEEF 418
QY 450 PFTIRVISEVLESNGSSSMASACGSLALMDSGVPISAVAGVALGLVTKTDPKGEIED 509
DB 419 PYTIRLVSEVLSNGSTSQASVCCSTLALMDAGVPIKRPAGIAMGLTSSDLKEAV-- 476
QY 510 YRLITLDIGIDVNDMDFKIAGTNKGITLQADIKLPGIPIKVMELIQASVAKKEIL 569
DB 477 ---ITDIOGLDPRGDMDFKVAEGTEKGITLQVDTKIHGSLSKYCIKTAINDARARLFI 533
QY 570 QIMNKTSIKPRASKRENGPVVETVOVPLSKRAKFFGPGVNLKTLQAEVTVISQVBET 629
DB 534 EKMAACINERPEKELSTYAPRAYTINIDDKRTILIGTGKTIINKIIEETGKILIREBGT 593
QY 630 FSVFAPSPVMEHARDPTEICDKDOEQLEFGAVYATATTEIRDTGVNVLVPMNTAVL 689
DB 594 VFVLSSDASDANRLAKMIDLTKD----VKVGEVYLGKVTIKINFGAFVEVLGKGS-L 647
QY 690 LAHTQOLDNERLN 701
DB 648 VHSIKLDINKVN 659

RESULT 6
US-09-815-242-5443

;; Sequence 5443, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haasbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zykend, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Grant, J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; TITLE OF INVENTION: Prokaryotes
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 1410
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 5443
;; LENGTH: 696
;; TYPE: PR1
;; ORGANISM: Staphylococcus aureus
US-09-815-242-5443

Query Match 33.5%; Score 1191.5; DB 9; Length 696;
Best Local Similarity 39.2%; Pred. No. 2.7e-84;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

QY 24 SAGRAVAVDLGNRKLKLEISSGKLARFADGSAVVOGPTAVMTAVSTKSPSPFMVLV 83
DB 2 SOEKVAVTKTWAGSLTIEFGOLAKQANGAVLVRYGDTVALSTATAKEPRDGFPLTV 61
QY 84 DYROKAAAGRIIPNVYLRREVSTDEKELTSRIIDRSIRPLPAGYFYDTQVLCNLAVD 143
DB 62 NYEKKMTAAAGKIPGVFKRGGRGDDATYTLARLIDRIRLFPKGYHDVQIMMVLSD 121
QY 144 GVNEPDVLAINGASVALSLSDIPMNGPVCAVRIGITDGYVNPTRKEMSSSTLNLVAG 203
DB 122 PDGSPQAAAMIGSSMALSVSDIPQGPVAGVNGYIDGKTIINPTVEKEVSRDLLEVAG 181
QY 204 APSQIVMLASAEINLQDFCHAIKVG-----VKYTOOIIQGLQOLYKENVYTRTP 256
DB 182 -HDAVVMVAVAGASEITEOEMLKAIFFGHEIORLVDFOQOIVDHIOVRQE----- 232
QY 257 QKLTSPS-----PEIVKTHKLAMERLYAVFTDYEHDKVSRDEAVNKRILDTESQLE 309
DB 233 ---FIPARDEBALVERIKSLTEKGLKEVYTF-----DIQOQDENIDNK---EELVNE 281
QY 310 KFEPRAD-----YEIESFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEDVMKTL 363
DB 282 FIDBEDENELLKEVYALINELVKEVRLIDDKIRPDGRKDEIRPLDSEVGILPRT 341
QY 364 HGSALFQRCQTOVLCTVTFPSLESIGKISDOVITAINIKIKKNPMLHFFPPYAINELGV 423
DB 342 HSGSLFTRQTOALSVTLTAL-----GDYQIIDLGLGPEBEKRFMHYHNFNFSVGETGPV 397
QY 424 TGNRRELGHALAELKALVPVP--RDPEFTIRVTSVLESNGSSSMASACGSLALMD 481
DB 398 RAPGRREIGHALGERALKYIIPDTADFPYTIIRIVSEVLESNGSSSQASIGSLALMDA 457

```
Qy 482 GVPISAVAGVATGLVTKTDPKEGEIEDYRLTLTDLIGIEDYNGMDFKIAGTNGKITALQ 541
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 GVPKAPVAGIAGMLVTRD-----SYTLTDIOGMEALGMDPFAVAKTGTITAIQ 510
Qy 542 ADILPGIPIKIVAEAIQOASVAKKEIIQINNTKTSKPRASRKENGVPVETVQVPLSKRA 601
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 MDIKIDLTREIIIEALQARGRLIEIMNMHLQITIDQRTLSAVAPKVTMTIKPKIR 570
Qy 602 KFPVPGVNLKQQAETGVTTISQVDEETFSVPAPTPSMHEARDPTEICKDDEQOLUF 661
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFGAVDQAMINRARIIEITR-----EAV 625
Qy 662 GAVYTATITEIRDTGVWVVKLYPNMTAVLHNTQDNERL 700
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 626 GQTYQATVKRIEKYGAVGLFPGKDA-LHHSQISKNRI 663

RESULT 7
US-09-815-242-12325
; Sequence 12325, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12325
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12325

Query Match 33.5%; Score 1191.5; DB 9; Length 698;
Best Local Similarity 39.2%; Pred. No. 2,76-84;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;
```

```
Qy 204 APRSQIWLSEASENIILQDFCHAIKVG-----VKYTOQIIIGIQQLVKTEGVTKRP 256
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 HKDAVMNVEAGASEIIEQEMLEAIFPGHEIIQGLVDFQOQIYVHIGPVKGE----- 232
Qy 257 QKLFPS-----PEIVKTYHKLAMERLYAVFTDYHDKYSRDAVANKIRLDTBEOLKE 309
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 ---FIPARDEALVERIKSLTEEGKLEKTVLTF-----DKQQRDENLDMNK---EELVNE 281
Qy 310 KFPBADP-----YEIIESFVNAKVEFPRISVLNMEYKRCQGRDLTSLRNVSCEVDMFTL 363
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 FIDEEDDENELIKEVAILNELVKEVRRLIADEKIRPDGRKDEIRLPDSEVIGLERT 341
Qy 364 HGSALFORGOTVLCCTYTPDSIESGISQDVITAINGIKDNFMALHYEFPYATNIGKV 423
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 HGSGLFTRGOTQOLSVLTIGAL-----GDYQILDGLGEBEERFHHNINFRFSVGEGPV 397
Qy 424 TGLNRRELGHGALAEKALYVPVIR--RDPEPTIRVTSVLEBSNGSSSMASACGSLALMDS 481
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 RAPRRRIIGHGALGERALKYIIPDTADFPYTIIRIVSEVLBSNGSSSOASICGSLTALMDA 457
Qy 482 GVPISAVAGVATGLVTKTDPKEGEIEDYRLTLTDLIGIEDYNGMDFKIAGTNGKITALQ 541
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 GVPKAPVAGIAGMLVTRD-----SYTLTDIOGMEALGMDPFAVAKTGTITAIQ 510
Qy 542 ADILPGIPIKIVAEAIQOASVAKKEIIQINNTKTSKPRASRKENGVPVETVQVPLSKRA 601
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 MDIKIDLTREIIIEALQARGRLIEIMNMHLQITIDQRTLSAVAPKVTMTIKPKIR 570
Qy 602 KFPVPGVNLKQQAETGVTTISQVDEETFSVPAPTPSMHEARDPTEICKDDEQOLUF 661
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFGAVDQAMINRARIIEITR-----EAV 625
Qy 662 GAVYTATITEIRDTGVWVVKLYPNMTAVLHNTQDNERL 700
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 626 GQTYQATVKRIEKYGAVGLFPGKDA-LHHSQISKNRI 663

RESULT 8
US-10-282-122A-44413
; Sequence 44413, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykend, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44413
; LENGTH: 698
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-10-282-122A-44413

Query Match      33.5%; Score 1190.5; DB 15; Length 698;
Best Local Similarity 39.1%; Pred. No. 3,3e-84;
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;

QY 24 SAGRAVAVDIGNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQFMPLV 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 SOEKRYFKTEWAGRSITIEGQLAKQANGAVLVRYGDTVLSTATASKERPDGDFPLTV 61

QY 84 DYQKAAAGRIPTNYLRREVGTSDEKELTSRIIDRSIRLPFPAGYFDPQVLCNLAVD 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 NYEKMTAAGKIPGKFKRBERPDATLTARLIDRPIRPLFPKGYKHDVQIMNVLSAD 121

QY 144 GVNEPDVLAINGASVALSLSDIPMNGFVAVRIGIIGDEVYVNPTRKEMSSSTLNVAG 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 PDGSPMAAMAGSMALSVSDIPFGPIAGVNVYIDGKVIINFTVEKEKSRDLDEAVG 181

QY 204 APSQIVMLEASABNTIQQDFCHAIKVG-----VKYTOQIIQGIQQLVKEGVTKRP 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 -HKDAVMVMEAGASBITESQMLEAIFGHEIRQLVDFQGIIVDHIQPVKE----- 232

QY 257 QKLFPS-----PEIVKYTHKLAMERLVAVFTDYHDKVSRDEAVNKIRLDTEQJKE 309
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 ---FIPARDEALVERVAKSLTEEGKLKETVLT-----DKQORDENLNLK---BEIVNE 281

QY 310 KFPKPAD-----YEIESFNVAKEVRSIVLNEYKRCDSRDLTSLRNVSCVDMFKTL 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 FIDEDEBENELLIEVVAIINELVKEVRRLADEKIRPDGRKDEIRLPDSEVGIILPRT 341

QY 364 HGSALFORGOTVLCVTVPDSLESIGKSDQVITAINGIKDKNFMILHYEPYPATNETIGKV 423
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 HGSALFRGOTVLCVTVPDSLESIGKSDQVITAINGIKDKNFMILHYEPYPATNETIGKV 397

QY 424 TGLNRRLHGALAEKALYPIR--RDPPTIRVTSVLENSGSSSMASACGSLAMD 481
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 RAPGRREIGHGALGERALKYIIPPTADFPYTIIRIVSEVLESNGSSSQASICGSLTAMD 457

QY 482 GVPISSAVAGVAGIYVTKTDEPEKEIEDYRLTDLILGIEDYNGMDFKIAGTKNGITALQ 541
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 GVPKAVVAGIYVTKTDEPEKEIEDYRLTDLILGIEDYNGMDFKIAGTKNGITALQ 510

QY 542 ADILPGLPIKTVMEAIQOASVAKKEIIQINNKTKISKRPRASRKNENGVPVETVQVLSRA 601
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 MDIKIDGLTREIIEEALQARGLRLEIMNHMLQTIIDPRTLSVAPVVMITIKPKIR 570

QY 602 KFPVPGGNLKKLAQFQVITISQVDEDEFVSFAFPTSPVMEHARDPFIETICDDDEQJLEF 661
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 DVIGPGGKKNELIIDEFGVKLDIFQDGTIFGAVDQAMINARBIIEITR-----EAEV 625

QY 662 GAVVTATITEIRDTGVWVKLYPNMTAVVLIHNTQDNEEL 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 626 GQTYQATVYKRIEKYGAIVGLFPGKDA-LIHISQISKRI 663

RESULT 9
US-10-282-122A-70948
; Sequence 70948, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70948
; LENGTH: 701
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70948

Query Match      32.8%; Score 1165.5; DB 15; Length 701;
Best Local Similarity 38.7%; Pred. No. 3e-82;
Matches 270; Conservative 123; Mismatches 250; Indels 55; Gaps 13;

QY 24 SAGRAVAVDIGNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQFMPLV 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 SOEKRYFKTEWAGRSITIEGQLAKQANGAVLVRYGDTVLSTATASKERPDGDFPLTV 61

QY 84 DYQKAAAGRIPTNYLRREVGTSDEKELTSRIIDRSIRLPFPAGYFDPQVLCNLAVD 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 NYEKMTAAGKIPGKFKRBERPDATLTARLIDRPIRPLFPKGYKHDVQIMNVLSAD 121

QY 144 GVNEPDVLAINGASVALSLSDIPMNGFVAVRIGIIGDEVYVNPTRKEMSSSTLNVAG 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 PDGSPMAAMAGSMALSVSDIPFGPIAGVNVYIDGKVIINFTVEKEKSRDLDEAVG 181

QY 204 APSQIVMLEASABNTIQQDFCHAIKVG-----VKYTOQIIQGIQQLVKEGVTKRP 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 -HKDAVMVMEAGASBITESQMLEAIFGHEIRQLVDFQGIIVDHIQPVKE----- 232

QY 257 QKLFPS-----PEIVKYTHKLAMER--LVAVFTDYHDKVSRDEAVNKIRLDTEQJKEF 311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 ---FVPERDEDLVEKYKSLTEDEKGLDVTLT---FDKQORDENLNLK---BEVGHFL 283

QY 312 PEADP-----YEIESFNVAKEVRSIVLNEYKRCDSRDLTSLRNVSCVDMFKTLHG 365
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 DEEDPENETLVKVVYALINDLIEKVRRLADEKIRPDGRKVDKIRLSEVGLLPRAHG 343

```


QY 366 SALFORQOTVLTCTVETDSSESIGKSDQVITAINIGIKDKNFMLYEPPVATNIEIGKVTG 425
DB 344 SGLFTRGOTQALSLTLTGAL----GDYQLIDGLPBEVEKRMHMYNPNFPGVETGPVRA 399
QY 426 LNRRLGHALAEALFVLP--RDPPFTIRVTEVLESNGSSSMASACGSLALMSGV 483
DB 400 PGRREIGHGALGERALRYIIPDTQDFPYTRIVSEVLESNGSSSQASICGSLTALMDAGV 459
QY 484 PISAVAGVAVIGLTKTDPKEGEIEDRYLLDILIGIDYNDMDPKIAGTKGITALOAD 543
DB 460 PIRAPVAGVAGLVTGTRD-----STLTIDIGMEBALGDMDFKAGTKGITALOMD 512
QY 544 IKLPGIRIKTWEAIIQASVAKKEILOIMNTISKPRASRENGPVVETVQVPLSKAPF 603
DB 513 IKIDGLREVIEBALAQGRALIMDMHTIEQPREELSAVAPKVTMSINDKIRDV 572
QY 604 VEGGKIVKLOAETGTTISQVDEETSVFAPTPSVHNEARDFTEICKDOEQLEPGA 663
DB 573 IGPGRKINEIIDEETGYKLDIEODGTIFIGAVDQAMINRAKEIIEDITR-----EAEVQ 627
QY 664 VYTATIRBDTGVVTKLYPMNTAVLHNTQDNERLN 701
DB 628 VTHAKVRIEKYGAFFELFEGKA-LHHSIQSERIN 664

RESULT 10

US-10-282-122A-71588
Sequence 71588, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forgyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 71588
LENGTH: 692
TYPE: PRT
ORGANISM: Staphylococcus haemolyticus

FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (470)..(470)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-71588

Query Match 32.7%; Score 1161.5; DB 15; Length 692;
Best Local Similarity 38.5%; Pred. No. 6.1e-82;
Matches 262; Conservative 126; Mismatches 247; Indels 45; Gaps 11;

QY 37 RKEIISGKARFADGSAVVOGPTAVMTAVSKTKSPQFPLVVDYKQKAAAGRI 96
DB 9 RSLTIEIGLAKQANGAVLVRYGDTVLSTAVASKEBRDDFPPLTYVEEKMYAAKIP 68
QY 97 TNYLRREVGSDKEILRSRIIDRSIRPLFPAGFYDQVLCNLAVDGNBPDVLAINGA 156
DB 69 GGFKRGRRGDEDTLRLIDRPIRLFPKGRHNDVQIMTYVASDPDGSPEKMAIGS 128
QY 157 SVALSLSDIPWNGVGAIRIGIDGEVYVNPTRKEMSSSTLNLVAGAPKSOIYMLASA 216
DB 129 SMLSVSDIPFGPIAGVNVGVYDGKVINPTVEKEKVSRLDEVAQ-HKDAVNVAGA 187
QY 217 ENLIQDPCHAIKVG-----VKYTOQIIIGIQQLVETGVTKRTPOKLFTESELYKY 269
DB 188 SEITEKEMLEAIFEGHDEIRQLVAFOREVVAHIQPVKEEFVPERDEALV-----SRVKT 243
QY 270 THKLAMERLVAVFTDYHDKVSRDEAVNKRIDTBEOLKEKPEADPY-----EIESF 323
DB 244 TEKGLKETVLT-----DKQQRDENLDTKAEIATEFVD--EADENELLIDEVAIL 295
QY 324 NVVAKEVFRSIVNVEYRCGRDLTSLRNVSCEDVMEKTLGSLFORQOTVLTCTVFD 383
DB 296 NDLYKEVRRLIADDEKIRPDGRKDEIRPLESEVGLIPRAHSGSLFRGQTQALSVTLIG 355
QY 384 SLBSGKSDQVITAINIGIKDKNFMLYEPPVATNIEIGKVTGLNRRRLGHALAEALY 443
DB 356 AL-----GDYQLIDGLGEGEKEKRFHMYNPNFPGVETGPVRAVGRREIGHGALGERALY 411
QY 444 VIP--RDPPFTIRVTEVLESNGSSSMASACGSLALMDSGVPLSSVAVAVAGLTKTD 501
DB 412 IIPDTTEFPYTVAVSVLESNGSSSQASICGSLTALMDGVPKAVVAIANGLVTRKD 471
QY 502 PEKEIEDRYLLDILIGIDYNDMDPKIAGTKGITALOADIKLPGIPIKIWEAIIQOA 561
DB 472 SNPD-----LTDIQMEBALGDMDFKAGTKGITALMDITREVEIEALQA 524
QY 562 SVAKKEILOIMNTISKPRASRENGPVVETVQVPLSKRAKFGVGGVNLKULOAGETVT 621
DB 525 RQGLAIIMDMHLCITIDQPRKELSAVAPKVEIMHIKPEKIRDIVGPGKKINEIIDEETGV 584
QY 622 ISQVDEETSVFAPTPSVHNEARDFTEICKDOEQLEFGAVYTAITTEIRDTGVNKL 681
DB 585 LDIEODGTIFIGIDQDINRAKEIIEDITR-----EAEVQYVNAKRIEKYGAFFEL 639
QY 682 YPMNTAVLHNTQDNERLN 701
DB 640 FAGKDA-LHHSIQSERIN 658

RESULT 11

US-10-282-122A-75529
Sequence 75529, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant

```

/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 75529
/ LENGTH: 711
/ TYPE: PRT
/ ORGANISM: Salmonella typhi
/ US-10-282-122A-75529

Query Match      32.6%; Score 1161; DB 15; Length 711;
Best local Similarity 39.9%; Pred. No. 6.9e-82;
Matches 270; Conservative 118; Mismatches 249; Indels 40; Gaps 11;

QY 35 GNRKLEISSGKLAFADGSANVOSGDTAVWVTAASKTKPSPQ-FMPLVVDYRQAKAAAG 93
DB 12 GQHTVLTLETGMMAQAPAAVAVWSMDTAVFVTVVGQKAKRGQDFPLVNVYQERTYAAG 71
QY 94 RIPTNYLRREVSGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLLAVDGVNEPDVLA 153
DB 72 RIPSFRRRGRSREGSTLRLRLDRVRLFPBGFVNEQVATVSVNQNVPDVLAM 131
QY 154 NGASVALSLSDIPWNGVAVRIGIIDGEYVNVPTRKEMSSSTLNLVVAAPKSQIYMLE 213
DB 132 IGASAAISSLGIPFNGPIGARVGVINDQYVIANPTODELAKESKLDLVVAGT-EAVALMVE 190
QY 214 ASAANIIOQDFCHAIKRGVVKTKQOIIQGIQOLVKEVTKRTPQKLPSPSEIYK----- 268
DB 191 SEAEILSEBDTMLGAVVGHQEQVVIQAINDLVKEAGPRADWG-----PEAVNDALINA 244
QY 269 YTHGLAMERT---YAVFTDYEDHKVSRDEAVNKIRLDTDEBOLKEKFPREADPEYELIESFN 325
DB 245 RVALAESRLSDAYRI-----TDQERYAQVDVAKSETIIEBLADEDTLDANEIGETILHA 299
QY 326 VAKEVFSIVINEYKRCGRDLTSLRVNSCEVDMFKTLHGSAIFQRCQTVQLCTVTPDSL 385
DB 300 IEKVIVSVRVLAGRPRIIDGREKMDIRGLDVRTGLPRTHGSAIPLTRBGTQALVATATLGT- 358
QY 386 ESGIKSDQVITAINGIDKXNMLHYEPPTATNETIGKTYGNRRELGHGALAELALPYI 445
DB 359 ---ARDQVAVDELNGERTDSFLFHYNPPYSGVGTGWGSKRREIGIRLAKRGVALVM 415
QY 446 P--RDPPTIRVTSEVLENGSSSMASACGSLALMDGVPISASVAGVALGLVTKPDPE 503
DB 416 PDMKFPYTVAVVEIITSNGSSSMASACGSLALMDGVPISASVAGVALGLVKEGD-- 473
QY 504 KGEIEDYRLLTDILGIEDYNGDMDFKLAGTNKGTALQADIKLPGIPIKIYMEAIQQA 563
```

```

DB 474 -----NVVLSIDLGBDHLGDMDFKAGSRDGLSALQNDIKLEGITKEIMQVLANQAKG 528
QY 564 AKKEIILQINKTKITSPKASHKENGPPVEYVQVPLSKRAKFGVGVNKKLQAEYGTIIS 623
DB 529 ARHLILGVMQALNAPRGDISFAPRHTIKISTDKIDVIGKGSVIRALVTEETGTIE 588
QY 624 QVDEBFVSFAPRPPSVHMERADPTITEICNDQOQLEFGCAVYATITIEDTGVMYKLYP 683
DB 589 IEDDGTIVKIAATGEEKAKVARIIRIEI-----PABIEVGRIYNSKVYRIYDFGAFVAGG 643
QY 684 NMTAVILLNHTQDLNERL 700
DB 644 KGEK-LVHISQIADKRV 659

RESULT 12
US-09-815-242-13768
/ Sequence 13768, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13768
/ LENGTH: 721
/ TYPE: PRT
/ ORGANISM: Salmonella typhi
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (1)...(721)
/ OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13768

Query Match      32.6%; Score 1158; DB 9; Length 721;
Best local Similarity 39.9%; Pred. No. 1.2e-81;
Matches 270; Conservative 117; Mismatches 250; Indels 40; Gaps 11;

QY 35 GNRKLEISSGKLAFADGSANVOSGDTAVWVTAASKTKPSPQ-FMPLVVDYRQAKAAAG 93
DB 22 GQHTVLTLETGMMAQAPAAVAVWSMDTAVFVTVVGQKAKRGQDFPLVNVYQERTYAAG 81
QY 94 RIPTNYLRREVSGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLLAVDGVNEPDVLA 153
DB 82 RIPSFRRRGRSREGSTLRLRLDRVRLFPBGFVNEQVATVSVNQNVPDVLAM 141
QY 154 NGASVALSLSDIPWNGVAVRIGIIDGEYVNVPTRKEMSSSTLNLVVAAPKSQIYMLE 213
```

```

Db 142 IGASALSLISGIFPNIGIGARVGVINDQVYLVNPTQBELKESKLDLVVAGT-EAAVLAMVE 200
Qy 214 ASAMENIQDPFCAIKVGVYKTOIIOGIGQVLVETGVTKTPKLTTPSPBYK----- 268
Db 201 SEABELSEDMLGVVGEHQOQVIAINDLVYKAKPRMDW-----FEAVNDALMA 254
Qy 269 YTHKLMERL---YAVTDVENDKVSDEAVNKRILTEBOLKPKPEADPYEIESFN 325
Db 255 RVALLAESRLSDAYRI-----TDKQERYAOVDYKSETIEOLIMEDETLDANEIGELIHA 309
Qy 326 VAKESPRSVILNBYKRCGRDLTSLRVNSCEVDMFKTLHGSALFQSGQVYLCVTEPDSL 385
Db 310 IENKVVASRLVAGPRIDGHEKMIKGLDVRTGLPRTHSALPRTBETQALVATLGT- 368
Qy 386 ESGIKSDQVYTAINGIKDKNFMHYEPPIATNEIGVYGLNRELGHALAEKALYPI 445
Db 369 ---ARDAQVLDLMGERTDLFFHYNPPYVSGETGVVSGPKREIGHGLAKRGVLA 425
Qy 446 P--RDPFTIRVSEVLENSGSSMASACGSLALMDSGVPISAVAGVAGIAGVTKTDP 503
Db 426 PDMKFFYVAVVSEITESSGSSMASVCGASLALMDAGVPIKAAVAGIAGVKEGD-- 483
Qy 504 KGEIEDYRLTLTGIEDYNGDMDFKTAGTNKGTALQADIKLPGIPIKIWMENIQOASV 563
Db 484 ----NYYVLSLIDGDEHDIGDMDFKVASGSDIGSALQMDIKIEGIRKEMQVVALNQAKG 538
Qy 564 AKKEILOINMKTISKPRASKEGPPVETVQVPLSKRAKRVGCGYNLKLQAEVYTI 623
Db 539 ARHLILVMEQAINAPRGDISFAPRIHTIKISTDKIKDVIKGGSGVIRALTEBTGTTIE 598
Qy 624 QVDEETSVFAPRPSVNHAEADFTIECKDQEQOOLFEGAVYATTEIETDGMVLYP 683
Db 599 IEDGYKIAATDEKAKVARIIEI-----TAEIVGRYNSKVRIVDPFGAFVAG 653
Qy 684 NMTAVLHNTOLDNERL 700
Db 654 GKGG-LVHISQIADKRV 669

```

```

RESULT 13
US-10-282-122A-61234
; Sequence 61234, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Fotsyryh, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282, 122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191, 078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206, 848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207, 727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230, 335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230, 347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242, 578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253, 625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257, 931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267, 636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 61234
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Legionella pneumophila
US-10-282-122A-61234

```

```

Query Match 32.5%; Score 1157; DB 15; Length 729;
Best Local Similarity 38.2%; Pred. No. 1.5e-81;
Matches 258; Conservative 134; Mismatches 244; Indels 40; Gaps 12;

```

```

Qy 35 GNRKLEISGKLARFADGSAVVOGDTAVWYTVASKTPEPS-QFMPLVVDYRQKAAAG 93
Db 11 GNRKLEISGKLARFADGSAVVOGDTAVWYTVASKTPEPS-QFMPLVVDYRQKAAAG 70
Qy 94 RPTNYLREVGTSDEKILTSRIIDRSIRLPAGYFDYQVLCNLLAVDGNBPVLA 153
Db 71 KIPGFRKRGSRSDNETLSRIIDRSIRLPAGYFDYQVLCNLLAVDGNBPVLA 110
Qy 154 NGASVALSLSDIPWNGVAVRIGIIDEVYVNPTRKEMSSTLNVLVAGAPKSOIYME 213
Db 131 IGASALSLISGIFPNIGIGARVGVINDQVYLVNPTQBELKESKLDLVVAGT-KDAILMVE 189
Qy 214 ASAMENIQDPFCAIKVGVYKTOIIOGIGQVLVETGVTKTPKLTTPSPBYK----- 266
Db 190 SEAOBLESIDIMRGAMLYGHEMMQVIXSIELRADVQSKR--PE---WKAPEIDTVLKAR 244
Qy 267 VKYTHKLMERLAVFTDHYEDKVSDEAVNKRILTEBOLKPKPEADPYEIESFN 326
Db 245 INDVAREVBAVLI-----KDKQRYQRLDELRBOTISLALNENDELADVLANNGEL 299
Qy 327 AKESFISVILNBYKRCGRDLTSLRVNSCEVDMFKTLHGSALFQSGQVYLCVTEPDSL 386
Db 300 ERSIVNRLIDGSRPRIDGRHRTVPRISIRTKLEBTHGSCLETRBETQALVATL---- 355
Qy 387 SGIKSD-QVITAINGIKDKNFMHYEPPIATNEIGVYGLNRELGHALAEKALYPI 445
Db 356 -GNERDAQIUDGISGESRDFMLHYNPPYVSGETGVVSGPKREIGHGLAKRGVLA 414
Qy 446 P--RDPFTIRVSEVLENSGSSMASACGSLALMDSGVPISAVAGVAGIAGVTKTDP 503
Db 415 PDANEPFYVAVVSEITESSGSSMASVCGASLALMDAGVPIKAAVAGIAGVKEGD-- 472
Qy 504 KGEIEDYRLTLTGIEDYNGDMDFKTAGTNKGTALQADIKLPGIPIKIWMENIQOASV 563
Db 473 ----RAVLTDLIGDEHDIGDMDFKVASGSDIGSALQMDIKIEGIRKEMQVVALNQAKG 527
Qy 564 AKKEILOINMKTISKPRASKEGPPVETVQVPLSKRAKRVGCGYNLKLQAEVYTI 623
Db 528 GRTHILGVNNALAEHRTLSQAPRTITMKVADKIRITIGGAGATIKLISSTGVSID 587
Qy 624 QVDEETSVFAPRPSVNHAEADFTIECKDQEQOOLFEGAVYATTEIETDGMVLYP 683
Db 588 IDSGVQLFSPRMALAEKQIKALI-----ABIEVGQTYGKSKIVDPGAFINL 642
Qy 684 NMTAVLHNTOLDNER 699
Db 643 GKGG-LVHISQIADKRV 657

```

```

RESULT 14
US-10-282-122A-53063
; Sequence 53063, Application US/10282122A

```

```
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeeldeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 53063
LENGTH: 703
TYPE: PRT
ORGANISM: Clostridium difficile
US-10-282-122A-53063

Query Match      32.5%; Score 1155; DB 15; Length 703;
Best Local Similarity 38.6%; Pred. No. 2e-81;
Matches 261; Conservative 136; Mismatches 249; Indels 30; Gaps 14;
```

```
Db 305 EEVRAKLIVHENVAPDRNKLKEIRPINCETGMPRAHSAFLTREGQTVLVNATLGLAGDV 364
Qy 389 IKSDQVITTAINGIKDNFMHYEPPEPYATNEIGKVTGLNRELGAGLAERKALVPIV 446
Db 365 QKLD---GLDEBENKRYMHYFPAYSVGEARPSGPRGRIGHGALERALLPVPISQ 420
Qy 447 RDPFPTITRVTSYLVESNGSSMASAGGSLALNDGVPPISSAVAGVALGLVTTDPKGE 506
Db 421 EEPFYAIRLVSEVLSNGSTSQASVCGSTLSLDAGVPIDMWAGIAMGLI-RHD--GK 476
Qy 507 IEDYRLITDILIGEDVNDMDFKIAGTNGKITLALQADIKLPGIPIKIVMEAIQOASVAK 566
Db 477 V---AVLSDIQGMEDHGDMDFKVAGTEYITAIQMDIKIDGIDKELQALQOAGRI 533
Qy 567 EIIQINMKTISKPRASKEGPEVETVOVPLSRKAFVGGYNNKKLQAEFTVTISQVD 626
Db 534 HILGEMKTIQPKPELSPYAKIVKQINPDKIKVIGGKILTKIIDIETGVKID--I 591
Qy 627 EEPFVAPFPPSV--MHEADPITEICKDQEQOQLERGANVYTAITIEIRDTGVNVLKPN 684
Db 592 EQGVEVFIGIEIDMTKKQDELNNIIVEP---EVGKTYKGVKSRIMFGAFVEILPG 646
Qy 685 MTAVLHNTQDNRL 700
Db 647 KEG-LHHSIAHERV 661

RESULT 15
US-09-815-242-10312
Sequence 10312, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haeeldeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
Prokaryotes
FILE REFERENCE: EITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10312
LENGTH: 734
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10312

Query Match      32.5%; Score 1155; DB 9; Length 734;
Best Local Similarity 39.9%; Pred. No. 2.2e-81;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;
```

```
Qy 35 GNRKLEISSGKLARFADGSAVVOGDPAAVNTAVSKTKPPSQ-FMPLVYVDYQKAAAG 93
```

Db 35 GQHTVLTLETGMMARQATAAVVMSMDTAVFTVVGQKKAKKPGODFFPLTVNYOERTYAG 94
QY 94 RIFPNVYARREVGTSDEKELTSRIIDRSIRLPFAGYFVDTOVLNCLAVDGNEDVLAI 153
Db 95 RIGSFFRRREGRPSEGETLRLIDRPIRLPPEGFVNEGVATVSVNPQVNDIVAM 154
QY 154 NGASVALSLSDIPNNGPVGAVRIGIIDGEVYNPFRKEMSSSTLNLVAGAPKSQIWMLE 213
Db 155 IGASALSLSGIPNNGPIGARVGIVNQVYLANPQDELKESKLDIVVAGT-EAVALMVE 213
QY 214 ASANILIQODFCHAIKVGVKTYQOIIGIQQLVKGVTKRPQKLFPPSP--EIVKYT 270
Db 214 SEAQLISEDQMLGAVFEGHEQQOVVIQINELVKEAG---KPRWDMQPEPVNEALNARV 269
QY 271 HKLAMERL---YAFPTYEHKVSDEAVNKIRLDTBEQLKEXPEADPVEIIESFNVA 327
Db 270 AALAEARISDAVRI---TDKQERYAQQVDVAKSETIATLAEDETLDENELGEILHAIE 324
QY 328 KEVRSIVLNEYKRCDSGLTSLNVSCEVDMFKTLHGSALFORGQTOVLCTVTFDSLES 387
Db 325 KNVRSRVLAGEPRIDREKXIMIGLDVRTGVLPRTGOSALFTRGETOALVTATLGT--- 381
QY 388 GIKSDQVITAINGIKDKNFMLEPPPYATNEIGKVTGLNRRRELGHGALAEKALYVIP- 446
Db 382 -ARDAQVLDELMGERTDTFLFHYNFPYXVGETGVGSPKRRRIGHGRLAKRGVLAVMPD 440
QY 447 -RDEPPTIRYTSVELESNGSSSMASACGSLALMDSGVPISSAVAGVALGLVTKTDPEKG 505
Db 441 MDKPYTVRVAVSEITESNGSSSMASVCGASIALMDAGVPIKAAVAAGIAMGLVKEGD--- 496
QY 506 EIEDYRLTDLIGIEDVNGMDPKIAGTNKGITLQADIKLPGIPIKIYMEAIQOASVAK 565
Db 497 ---NYVLSDILGDEBDHIGMDPKVAGSRDGISALQMDIKIEGITKEIMOVALNQAQGAR 553
QY 566 KEILOINUKTISKPRASRKENGPEVETVQVPLSKRAKFPVPGYNLKKLQAEVTGISQV 625
Db 554 LHIGVMEQAINABRGDISBPAPRIHTIKINPDKIKOVIGKGGSVIRALTEETGTTEIE 613
QY 626 DEETFSVPAPTPSVMEARDFITRICKDDOQOLEFGAVTYATITEIRDTGVMTKLYPNM 685
Db 614 DDGTVKIAATDEKAKGAIIRIEI---TAIEIVGGRVYGVKTRIVDFGAFAVIGGK 668
QY 686 TAVLHNTQLDNERL 700
Db 669 EG-LVHISQIADKRV 682

Search completed: January 28, 2005, 19:47:36
Job time : 150 secs

This Page Blank (uspto)